Aix-Marseille Université

Diffusion des épidémies: le role de la mobilité des agents et des réseaux de transport

Epidemic spreading: the role of host mobility and transportation networks

Thèse présentée par

Paolo Bajardi

pour obtenir le grade de

Docteur d'Aix-Marseille Université

Spécialité: Physique Théorique et Mathématique
Ecole Doctorale de rattachement: Physique et Sciences de la Matière

À soutenir le 24 Novembre 2011 devant le jury composé de:

A. Barrat	Directeur de Thése
F. Carrat	Rapporteur
V. Colizza	Co-directrice de Thése
P. Jensen	Rapporteur
R. Pastor-Satorras	Membre du Jury
J.-F. Pinton	Membre du Jury

Résumé

Ces dernières années, la puissance croissante des ordinateurs a permis à la fois de rassembler une quantité sans précédent de données décrivant la société moderne et d'envisager des outils numériques capables de s'attaquer à l'analyse et la modélisation les processus dynamiques qui se déroulent dans cette réalité complexe. Dans cette perspective, l'approche quantitative de la physique est un des catalyseurs de la croissance de nouveaux domaines interdisciplinaires visant à la compréhension des systèmes complexes techno-sociaux. Dans cette thèse, nous présentons dans cette thèse un cadre théorique et numérique pour simuler des épidémies de maladies infectieuses émergentes dans des contextes réalistes. Dans ce but, nous utilisons le rôle crucial de la mobilité des agents dans la diffusion des maladies infectieuses et nous nous appuyons sur l'étude des réseaux complexes pour gérer les ensembles de données à grande échelle décrivant les interconnexions de la population mondiale. En particulier, nous abordons deux différents problèmes de santé publique. Tout d'abord, nous considérons la propagation d'une épidémie au niveau mondial, et présentons un modèle de mobilité (GLEAM) conçu pour simuler la propagation d'une maladie de type grippal à l'échelle globale, en intégrant des données réelles de mobilité dans le monde entier. La dernière pandémie de grippe H1N1 2009 a démontré la nécessité de modèles mathématiques pour fournir des prévisions épidémiques et évaluer l'efficacité des politiques d'interventions. Dans cette perspective, nous présentons les résultats obtenus en temps réel pendant le déroulement de l'épidémie, ainsi qu'une analyse a posteriori portant sur les stratégies de lutte et sur la validation du modèle. Le deuxième problème que nous abordons est lié à la propagation de l'épidémie sur des systèmes en réseau dépendant du temps. En particulier, nous analysons des données décrivant les mouvements du bétail en Italie afin de caractériser les corrélations temporelles et les propriétés statistiques qui régissent ce système. Nous étudions ensuite la propagation d'une maladie infectieuse, en vue de caractériser la vulnérabilité du système et de concevoir des stratégies de contrôle. Ce travail est une approche interdisciplinaire qui combine les techniques de la physique statistique et de l'analyse des systèmes complexes dans le contexte de la mobilité des agents et de l'épidémiologie numérique.

Abstract

In recent years, the increasing availability of computer power has enabled both to gather an unprecedented amount of data depicting the global interconnections of the modern society and to envision computational tools able to tackle the analysis and the modeling of dynamical processes unfolding on such a complex reality. In this perspective, the quantitative approach of Physics is catalyzing the growth of new interdisciplinary fields aimed at the understanding of complex techno-socio-ecological systems. By recognizing the crucial role of host mobility in the dissemination of infectious diseases and by leveraging on a network science approach to handle the large scale datasets describing the global interconnectivity, in this thesis we present a theoretical and computational framework to simulate epidemics of emerging infectious diseases in real settings. In particular we will tackle two different public health related issues. First, we present a Global Epidemic and Mobility model (GLEaM) that is designed to simulate the spreading of an influenza-like illness at the global scale integrating real world-wide mobility data. The 2009 H1N1 pandemic demonstrated the need of mathematical models to provide epidemic forecasts and to assess the effectiveness of different intervention policies. In this perspective we present the results achieved in real time during the unfolding of the epidemic and *a posteriori* analysis on travel related mitigation strategies and model validation. The second problem that we address is related to the epidemic spreading on evolving networked systems. In particular we analyze a detailed dataset of livestock movements in order to characterize the temporal correlations and the statistical properties governing the system. We then study an infectious disease spreading, in order to characterize the vulnerability of the system and to design novel control strategies. This work is an interdisciplinary approach that merges statistical physics techniques, complex and multiscale system analysis in the context of hosts mobility and computational epidemiology.

Contents

1

Introduction

Since the mid-20th century, the availability of new computational resources has added novel and important tools to theoretical physics. In particular, statistical mechanics had found in computer simulations the perfect environment to perform numerical experiments to explore ideal systems composed by thousands of interacting elements. Nowadays, the technological innovations have pushed the computational power to an extent unthinkable even twenty years ago. The easiness to which it is possible to perform huge calculations on ordinary personal computers represents a precious opportunity to tackle the study of natural processes involving a multitude of interacting objects in a complex reality, shifting the attention from ideal systems to the real-world phenomena.

Along with the technical innovations, the last decade has witnessed an increasing attention to the data gathering, ranging from biological (1; 2; 3; 4) to social systems (5; 6; 7; 8; 9; 10; 11; 12), from infrastructural (7; 13; 14; 15; 16) to financial realities (17; 18; 19; 20). In particular, the pervasiveness of technology in the every day life allowed physicists to quantitatively explore new fields. For instance, the mobile phones provide informations about the geographical position of his/her owner (21) and can be used as a proxy of human mobility as well as the tracking of banknotes (22). Furthermore, the analysis of mobile communication (23) along with the web 2.0 and the virtual social networks (24) permit the analysis of social interactions at a very large scale considering the interactions among thousand of individuals and measuring the conjectures and the hypothesis proposed by sociologists (25) and anthropologists (26) since the last century. Moreover, research on some technological systems pointed out their tendency to evolve as autonomous systems even though they have been built in their constituent elements by human beings, as in the case of the Internet (27) or the World-Wide-Web (28). All the above

mentioned systems are composed by many interacting entities and in order to formalize such interactions and to conveniently handle them, many datasets have been naturally described in terms of networks. The analysis of such systems highlighted their intrinsically complex nature: their behavior cannot be described extrapolating the properties and the governing laws of their constitutive units. The study of each subpart of the system in isolation does not allow the understanding of the whole system and its dynamics. In many cases, we observe the spontaneous outcomes of emerging phenomena due to the interactions among the system constitutive units. They are self-organizing systems without a blueprint or a global supervision.

The ambitious aim of the new computational science is to understand and predict through a data-driven approach the behavior of large-scale phenomena of the above mentioned techno-social systems. Starting with the mathematical description of patterns found in real data, scientists can devise models to anticipate trends, in order to evaluate risks and eventually manage future events (29). The most successful example of the potential of the computational modeling approach to complex settings is represented by weather forecasts: powerful supercomputers elaborate current meteorological data and correlate with huge libraries of historical entries into large-scale computational simulations achieving extraordinary results. Using the physical laws governing the dynamics of fluid and gas masses and sophisticated equipment to record data at the local level, in the last decades the development of accurate weather forecasts permitted to project the path and intensity of storms, hurricanes, and other disruptive meteorological occurrences and, in many cases, to save thousand of lives by anticipating and preparing for these events. On the other hand, the main obstacle for the understanding of human-related phenomena was the lack of large-scale data about human patterns and consequently the difficulty in formalizing mathematical laws governing human behavior. With the increasing availability of data about the individual and collective human dynamics, we are finally able to study phenomena involving social systems with the quantitative approach of statistical Physics. In this perspective, the modeling of epidemic spreading of infectious diseases is a challenging problem with immediate and important applications related to public health issues. Similarly to weather forecasts, we have to feed with the appropriate data and initial conditions the set of laws governing the transmission of pathogens in a spatially extended systems. The dissemination of infectious diseases is governed by host mobility at different scales. For instance, a child that gets influenza interacting with schoolmates can transmit the disease to her/his parents when s/he comes back home. Subsequently the parents might infect their colleagues at the workplace that is eventually placed in a neighboring town and the infection chain might continue infecting more and more people. In the pre-industrial

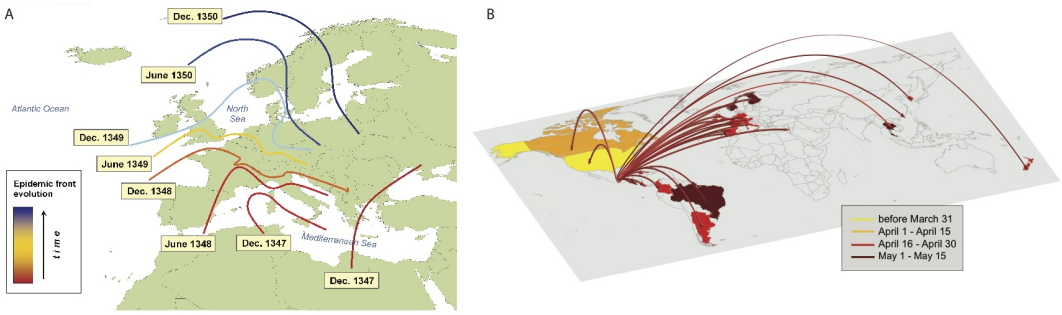

Figure 1.1: Disease spreading examples. The slow propagation of the black death in the 14th century (A) is compared with the rapid dissemination at the global scale of the 2009 H1N1 influenza (B).

age the human mobility was characterized by rare travels covering rather short distances, leading to a slow dissemination of new pathogens. As shown in figure 1.1A, it has been estimated that in the 14th century the so-called 'Black Death' was able to spread at the velocity of few hundreds miles per year. On the contrary, the mobility in the modern society is shaped by the large fluxes of short range daily commuters and air travelers. In such settings the spreading of infectious diseases may rapidly reach global proportions as in the paradigmatic case of the 2009 H1N1 pandemic shown in figure 1.1B. In this thesis, integrating with a network science approach the large-scale datasets describing the host mobility, we present a theoretical and computational framework to simulate epidemics of emerging infectious diseases in real settings.

The thesis is organized as follow. In chapter 2 we recall the basic concepts of graph theory, we introduce some mathematical tools that are necessary to analyze networked systems and we review three models that represent milestones of network science and are useful to understand the behavior of complex networks. In chapter 3 we provide an introduction to mathematical modeling of infectious diseases. The transmission of a pathogen within a host population can be described as a dynamical process governed by a set of differential equations by defining an appropriate phase space. In chapter 4 we describe in detail the Global Epidemic Mobility model (GLEaM) that integrates a data-driven network description of human mobility and an epidemic model of an influenza like illness. The work concerning GLEaM has been carried out within a large collaboration led by Vittoria Colizza (PI) and Alessandro Vespignani (PI and Team Coordinator) based mainly at the I.S.I Foundation, Turin, IT and at the Indiana University, Bloomington, IN, USA. Under the supervision of Vittoria Colizza, I have partially helped in the data gathering and

analysis of short range mobility datasets, while the computational implementation of GLEaM and the numerical simulations of epidemic scenarios has been an important part of the work performed during this thesis and it represents the crossing point of the interdisciplinary merging of human disease modeling, complex networks science and policy decision making. In chapter 5 we present the results obtained by using GLEaM to perform forecasts, analysis and intervention assessments during the 2009 H1N1 pandemic influenza. It is worth to stress that most of those results were achieved well before the epidemic peak opening the road to real time epidemic forecasting. In chapter 6 we present the results achieved under the supervision of Alain Barrat and Vittoria Colizza aimed at the development of novel mathematical and statistical tools through the longitudinal analysis of a dynamical complex network. Many systems are usually treated as static networks mainly because of a lack of information on the timing of the interactions and also because the static description is often a good approximation for several purposes. Nevertheless, when the time scale of the system dynamics is comparable to the investigated dynamical process it is crucial to incorporate the whole temporal information. We thus present the study on a detailed dataset of livestock movements that represents a unique opportunity to define new quantities and physical observables to analyze networked systems with an explicit temporal dimension. Finally, we further discuss some important features that may affect the spreading of an emerging infectious disease and we propose a novel approach to study dynamical processes on dynamical networks.

2

Theoretical Framework: Networks and Graphs

Contents

The first scientist to introduce the notion of graph was Leonard Euler in the famous work *Solutio problematis ad geometriam situs pertinentis* in 1736, where he solved the Königsberg bridges problem.

Generally speaking, a *graph* is an abstract way of specifying relationships among a collection of objects. Such level of abstraction can be applied to a broad range of systems. In this perspective graphs provide a theoretical framework that allows a convenient conceptual representation of interrelations in complex systems where the system characterization implies the mapping of interactions among a large number of constituent elements. The study of graphs has a long tradition in discrete mathematics, sociology, and communication research and has recently become very popular also in physics and biology.

In this work, the concept of networks plays a crucial role providing a suitable framework to explore computationally the unfolding of dynamical processes in complex realities and to tackle

theoretical issues raised by the dynamics of the system itself. In this chapter we will introduce the very basic concepts of this field but we refer the reader for a deeper analysis of the subject to some classical books for a more theoretical point of view (30; 31; 32; 33; 34; 35; 36) and for a more applied perspective (27; 37; 38; 39; 40; 41).

2.1 Basic definitions

A graph $G(N, E)$ is identified by a set of N objects named *vertices* or *nodes* and a list of E pairs of nodes, called *edges* or *links* indicating the relationships between the objects. Two nodes are *neighbors* if they are connected by an edge. The relationships between the objects can be symmetric or asymmetric leading to *undirected* or *directed* graphs respectively. Graphs are useful because they serve as mathematical models of network structures and from now on we will refer to graphs or networks without further distinctions. A convenient way to mathematically describe a graph is through the $N \times N$ *adjacency matrix* A, whose element $A_{ij} = 1$ if the nodes i and j are connected and $A_{ij} = 0$ otherwise. Using such formalism, the undirected networks are represented by a symmetric adjacency matrix ($A_{ij} = A_{ji}$), while in the directed cases the matrix can be asymmetric. Moreover, in the following we will consider only networks without self-loops ($A_{ii} = 0$). Depending on the system under study, it might be important to add a new degree of freedom representing the intensity of the relationships between nodes. It is thus possible to construct *weighted networks* where each connection (i, j) has its own weight w_{ij}.
In the following we recall some fundamental concepts and definitions used for the quantitative analysis of complex networks.

Path, connectivity and distance

A path P_{ij} defined in graph $G(N, E)$ is an ordered collection of edges connecting the nodes i and j. A graph is called *connected* if for every pair of nodes there is a path between them. A *component* C is defined as a connected subgraph and two components $C_1(M_1, E_1)$ and $C_2(M_2, E_2)$ are disconnected if it is impossible to construct a path P_{ij} with $i \in M_1$ and $j \in M_2$. Graphs usually lack a metric, but the distance between two vertex can be naturally defined as the number of links traversed by the shortest connecting path. For two given nodes i and j the *distance* l_{ij} between them is the path length with the minimum number of links between them. If the nodes i and j belong to two disconnected components, their distance is set to infinity. The *diameter* D of a graph is the maximal distance among all the pairs of nodes.

The average distance L is evaluated averaging the shortest paths for all the possible couples of nodes (i, j):

$$L = \frac{2}{N(N-1)} \sum_{i<j} l_{ij} \tag{2.1}$$

If L is much smaller than the size of the system N, the network is said to exhibit a *small world property*.

Degree

The *degree* k_i of a node i is the number of links connected to i, i.e. the number of its neighbors. By using the adjacency matrix formalism $k_i = \sum_j A_{ij}$ while the average degree is simply $\langle k \rangle = 2E/N$. In a directed graph the number of incoming or outgoing connections are called in-degree k^{in} or out-degree k^{out} respectively. The definition of degree can be generalized to weighted graphs where the weighted degree is usually called *strength*. The strength a the node i is $s_i = \sum_j w_{ij}$, and similarly to the unweighted directed graphs it is possible to define the in-strength and the out-strength quantities.

Centrality measures

When considering a network, the centrality of the nodes and links is crucial to understand their role in the systems. The degree is one of the simplest and more common centrality measure adopted to assess the node centrality, but many others exists: *betweenness* node/link centrality, *closeness* centrality, *eigenvector* centrality, *pagerank*. We will not go into details of such measures since in this work we will use the degree whenever we investigate the centrality of a node and we refer the reader to classical textbooks (37; 38; 39; 40) to deepen this topic.

Clustering coefficient

The clustering coefficient C_i of a node i is the fraction of neighboring nodes that are also connected to each other. In social networks the clustering coefficient quantifies the abundance of triadic closure, counting the prevalence of friends of a node that are also friend to each other. Let us consider a vertex i with $k_i = 3$ and let us imagine that two of them are connected with each other. In this case $C_i = 1/3$ because just 1 pair among 3 is actually connected. If all the neighbors were connected the clustering coefficient would have the maximum value $C_i = 1$. Using the adjacency matrix we have:

$$C_i = \frac{2}{k_i(k_i-1)} \sum_{j,k} A_{ij} A_{ik} A_{jk} \tag{2.2}$$

Degree distribution

The degree distribution $P(k)$ of a network represents the probability that a randomly chosen node has degree k. The average degree $\langle k \rangle$ is:

$$\langle k \rangle = \sum_k kP(k) \equiv \frac{2E}{N}. \tag{2.3}$$

A graph is called *sparse* if the average degree is very small with respect to the number of nodes: $\langle k \rangle << N$. In the case of a directed graph we have two distributions $P(k_{in})$ for the in-degree and $P(k_{out})$ for the out-degree. In the next sections we will present different classes of networks and we will discuss how the degree distribution is crucial to characterize a graph.

Degree correlation

Many real networks show a correlation between the degree of a node and the degree of its neighbors e.g. nodes with high degree are preferentially connected with high degree nodes. This kind of correlation is called *assortative mixing* and it is particularly common in social networks. In other cases the opposite situation has been found i.e. high degree nodes connected preferentially with low degree nodes. This kind of correlation is called *disassortative mixing* (42) and is rather common in technological and biological networks. Formally these correlations can be measured considering the average degree of the nearest neighbors of a generic node i, $k_{nn,i}$:

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \nu_i} k_j \tag{2.4}$$

where the sum is over the nearest neighbors of i. From this quantity a convenient measure to investigate the behavior of the degree correlation function is obtained by the average degree of the nearest neighbors, $k_{nn}(k)$ of nodes of degree k. This quantity can be expressed as:

$$k_{nn}(k) = \sum_{k'} k' P(k'|k) \tag{2.5}$$

where $P(k'|k)$ is the conditional probability that any given edge departing from a node of degree k is pointing to a node of degree k'. When the degrees of neighboring vertices are uncorrelated, $P(k'|k)$ is only a function of k' and thus $k_{nn}(k)$ is constant. In the presence of correlations $k_{nn}(k)$ could be a non constant function of the degree and in particular a positive (or negative) correlation is a signature of assortativeness (or disassortativeness).

2.2 Real networks

Much research has been done in the analysis of networked systems available from empirical datasets. In this section we will give some examples of real world networks. We will focus in particular on the macro areas of social, technological and biological networks including transportation infrastructures, human communication and mobility patterns (3; 14; 15; 16; 21; 22; 43; 44; 45; 46; 47; 48; 49). Studies performed on such different fields have unveiled the presence of unexpectedly similar properties, shared by these systems independently of their function, origin and scope. Besides the small world property, which consists in the co-existence of high local interconnectedness and small distances across any two nodes in the network compared to the system size (7), the components of such systems are found to be wired in a non-homogeneous way, with the number of connections per node showing very large fluctuations in contrast with the random Poissonian hypothesis (8). The ubiquitous nature of this so-called scale-free property - found across natural, societal, and artificial systems - has spurred more than a decade of research aimed at characterizing and understanding complex systems drawn from different disciplines through the common paradigm of networks science (50).

2.2.1 Social networks

Social Networks represent the individuals as nodes and the social interactions among them (friendship, sexual relations, belonging to the same group of work) as links. This kind of networks has been studied since the works of Moreno (51) in 1934 and are extremely important not just for social sciences but even for a wide variety of processes from the spreading of infectious diseases to the emergence of consensus and knowledge diffusion. The historical problem related to these networks was the difficulty to get reliable information of a sufficiently large number of people in order to have enough statistical power. Fortunately the recent explosion of online social interactions has made available data sets of unprecedented size. E-mail exchanges (5; 6), habits and shared interest inferred from web visits and professional communities such as collaboration networks of film actors (7; 9; 52) or company directors networks (53) or co-authorship among scientists (10; 11; 12) are classical examples of this type of networks.

2.2.2 Technological networks

Technological networks are human-built networks designed to accomplish the distribution of some resource: water, electricity, gas etc.. Classical examples are: the networks of power grids

both high or low voltage (7; 13), the networks of inter-urban streets (14), internet (27; 54; 55) and the airport networks (15; 16). This last system can be represented as a weighted graph where nodes are the airports, links the air connections and weights the flow of passengers. For more details we refer the reader to the Chapter 4 were a complete description of such network is presented. Another important technological network often classified as an information network is the *WolrdWideWeb*. It is the most famous virtual network where nodes are web pages and links are the hyper-links (direct links) between them. The extremely rapid and unregulated growth of the Web has led to a huge complex network. Its structure is very difficult to study and for many years experiments have been done in order to get information about it (28; 56).

2.2.3 Biological networks

Biological networks completely pervade the biological world spanning from the microscopic realm of biological chemistry, genetics, proteomics to the large scale of food webs. An important example is the protein interaction network (PIN) of various organisms where nodes represent proteins and edges connect pairs of interacting proteins (1; 2). Three different scales of processes are usually considered. The microscopic scale such as PIN networks in which the main point is to understand the biological significance of the topology of these networks (3). At a larger scale biological networks can describe interactions between animals and even humans (57). At the very large scale we find the networks describing the food webs of entire ecosystems (4).

2.3 Network models

The study of networked systems can be tackled with two complementary approaches. The characterization of real datasets and the analysis of different case studies has to be integrated with the development of models aimed at investigating the aggregation mechanism behind the observed patterns. In this perspective, many models have been formulated to explain and recover the main characteristics of the observed empirical networks. In this section we will recall just three of them, that for historical reasons and for the importance of their results are probably the most cited works in the field of complex networks.

2.3.1 Random networks: Erdös-Rényi (ER) model

The main contribution to the study of random graphs are due to Paul Erdös and Alfréd Rényi (58; 59; 60). In their first work they defined a random graph of N vertices and m links selected

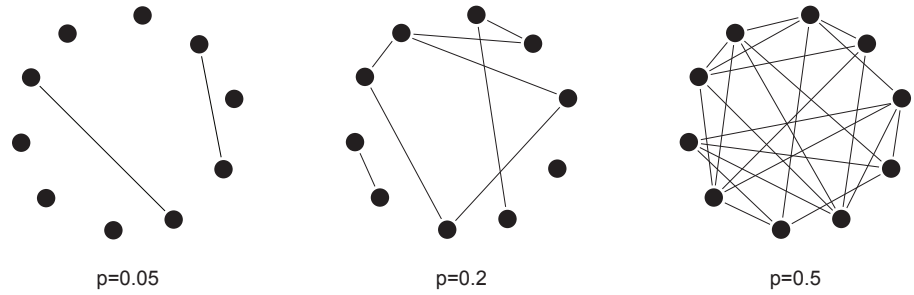

Figure 2.1: Schematic illustration of ER model with different connectivity probability. The larger is p, the denser is the network. When $p = 0$ the graph is composed by isolated nodes and when $p = 1$ the graph is fully connected.

at random among the $N(N-1)/2$ pairs of nodes. There are in total $\binom{N(N-1)/2}{m}$ possible graphs. They can appear with the same probability and they form the ensemble of graphs characterized by this rule. Another definition of random graph is given by the binomial model that is equivalent to the ER model for large N. Starting from N vertices for each pair of nodes a link is formed with probability p as illustrated in figure 2.1. The number of links is then a random variable with average value $\langle m \rangle = pN(N-1)/2$.

In a random graph characterized by a probability of connection p, in the limit of $N \to \infty$, the degree distribution can be approximated by a Poisson distribution:

$$P_{rand}(k) \simeq e^{\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{2.6}$$

The most characteristic trait of the degree distribution of random graphs is that it decays exponentially for large values of k allowing only very small degree fluctuations. The degree of the different nodes can thus be considered as uniform and equal to the average degree $k \simeq \langle k \rangle \simeq pN$. In general, random graphs are characterized by very small diameters showing a *small-world* behavior that is observed in many real world networks. It is easy to show that the diameter D is proportional to $ln(n)/ln(\langle k \rangle)$ (61). Moreover, random graphs are characterized by very small clustering coefficients. Given a node i, the probability that two of its neighbors are connected is equal to the probability that any other two nodes will be connected, so that:

$$C_{rand} = p = \frac{\langle k \rangle}{N}. \tag{2.7}$$

This implies that the ratio $C_{rand}/\langle k \rangle$ at fixed $\langle k \rangle$ decrease with the size of the system like N^{-1}.

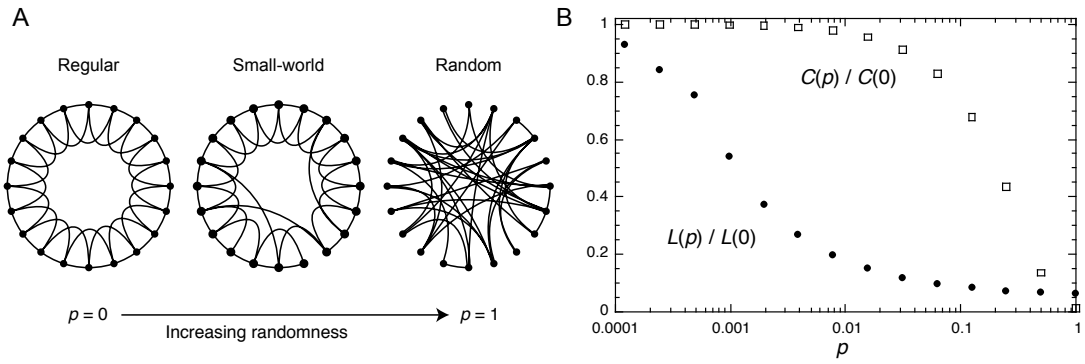

Figure 2.2: (A) The Watts-Strogatz model interpolates between a regular ring lattice and a random network with the random rewiring procedure. (B) Average shortest path length L and average clustering coefficient as a function of p. The quantities have been normalized by their value for the regular lattice topology. The rapid drop of L triggers the onset of small-world effects and it occurs when the clustering coefficient is still high.

2.3.2 Small world networks: Watts-Strogatz (WS) model

It has been observed that many real networks exhibit many closed triads leading to large clustering coefficients (like some regular graphs but in contrast with random graphs), but also very short paths (like random networks but in contrast with regular graphs). From this simple observation we can conclude that real networks are neither regular lattice nor random graphs, and following this inspiration, in 1998, Watts and Strogatz put forward a model to interpolate between these two limits (7). The model states that starting from N nodes in a ring, each connected to $k/2$ to the left and $k/2$ nodes to the right, every link is randomly rewired with a probability p as shown in Figure 2.2A. With this process $pNk/2$ links will be reshuffled on average. Nodes that before were far away from each other by construction are now closer thanks to the presence of shortcuts. It is extremely interesting to study the behavior of the average path length L of the graph and of the clustering coefficient C as a function of p. For $p = 0$ we have a complete regular ring with $L(0) \simeq \frac{n}{2k}$ and $C(0) \simeq \frac{3}{4}$ while for $p = 1$ the network becomes completely random with $L(1) \simeq \frac{ln(n)}{ln(k)}$ and $C(1) \simeq \frac{k}{n}$. For intermediate values of p we will have a situation in between these two limits. As shown in Figure 2.2B as p increases the distance gets reduced a lot while the average clustering is almost constant. The two quantities change with a complete different slope with p and there is thus a broad region in which we see a small-world effect and a value of clustering bigger than the random case.

It has been shown in Ref. (62) that the shape of the degree distribution for $p > 0$ is similar to the distribution of a random graph with a well defined peak around the mean value and an exponential decay for large value of k. This model is able to create small-world networks highly clustered but it still has a homogeneous topology and the degree distribution is far from being heavy-tailed as observed in real networks.

2.3.3 Scale free networks: Barabási-Albert (BA) model

In the last ten years many models have been proposed to explain the abundance in many natural systems of networks with high heterogeneity in the degree distribution. In the following we will recall the first, a very elegant, simple and most cited model, proposed in 1999 by Barabási and Albert able to produce scale-free graphs with small-world phenomena (8; 63).
In contrast with the previous models where the number of nodes N were fixed *a priori* and the edges were drawn (or rewired) with a certain probability, the BA model suggests a reasonable mechanism of networks formation. Here the nodes are progressively added to the system connecting with new edges the most popular links. Starting from a small number of core nodes n_0 the graph is thus build following these rules:

1. *growth*: at each time step a new vertex is added to the graph and it is linked with $m < n_0$ other nodes already present

2. *preferential attachment*: the new node is connected to the node i with probability π depending on the degree k_i
$$\pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{2.8}$$

Numerical simulations and analytical calculations show that the degree distribution of such networks will be:
$$P(k) \simeq k^{-\gamma_{BA}} \qquad with \quad \gamma_{BA} = 3. \tag{2.9}$$
It has been shown in (64) that the diameter of a BA graph is smaller than the relative measure of a random graph. The diameter of a BA graph is
$$D \sim \frac{\ln N}{\ln \ln(N)}, \tag{2.10}$$
It is then possible to show that a BA graph has a clustering coefficient higher than the random graphs of the same size and generated with the same $\langle k \rangle$ (65)(66)(67). The BA model

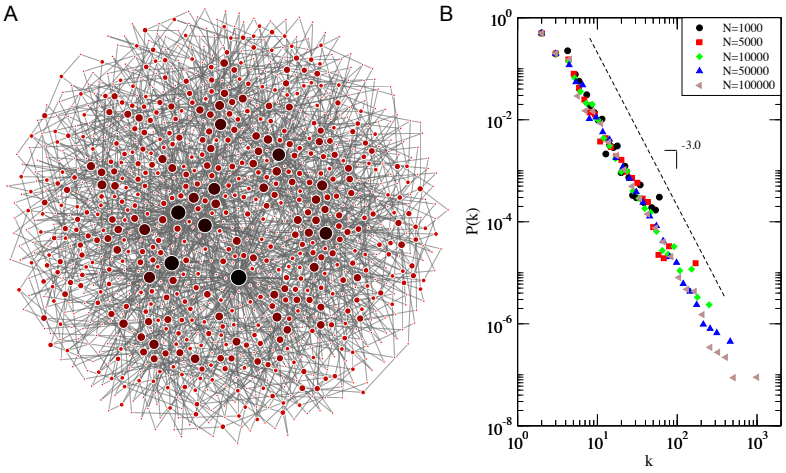

Figure 2.3: (A) Network with $N = 1000$ nodes generated with a BA model. The size and the color gradient of the nodes is proportional to their degree. (B) Degree distributions of networks generated with the BA model and different sizes.

has stimulated a lot of research in modeling the aggregating mechanism leading to networked structures. Nowadays there is abundance of models with tunable parameters able to create scale free networks. In particular with such refined models it is possible to tune the slope of the degree distribution, the clustering coefficient and the degree correlations.

The models presented in the last section will be used as benchmarks or paradigmatic examples to discuss the structural properties of the networks that will be described in the following chapters. It is worth to remember that the functional form of the statistical distributions characterizing large-scale networks defines two broad network classes. The first refers to the homogeneous networks where the degree distribution has a fast exponentially decay. The second class concerns networks with statistically heterogeneous connectivity patterns usually corresponding to skewed and heavy tailed distributions that can be approximated by a power law decay $P(k) \simeq k^{-\gamma}$. In general, in the asymptotic limit of $N \to \infty$, the cut-off k_c corresponding to the largest possible degree value diverges, so that $\langle k^2 \rangle \sim \int k^2 P(k) dk \sim k_c^{3-\gamma} \to \infty$: for values of $\gamma < 3$ fluctuations are unbounded and depend on system size. The absence of any intrinsic scale for the fluctuations implies that the average value is not a characteristic scale for the system. Whenever a network exhibits such large degree heterogeneities it is called *scale-free* and with the *small-world* effect it is one of the principal characteristics of *complex networks*.

2.4 Dynamical networks

As briefly enumerated in the previous section, many real systems can be treated as networks by coupling the constituent elements, namely the vertices, through some functional connections, namely the links. Nevertheless, in some cases, edges are active only for a certain period of time: e.g., a sexual contact between two individuals occurring at given time holds for a limited period (68). Similarly, a social interactions of people attending a conference can be represented by a graph where an edge between two individuals is on throughout the time they are chatting or at least in a close proximity (69). Like network topology, the temporal structure of edge activations can affect dynamics of systems interacting through the network, from disease contagion to information diffusion. The emergent field of temporal networks is exponentially growing and it is still lacking a unified theoretical framework to analyze the temporal datasets. In the light of traditional network theory, one can see this framework as moving the information of *when* things happen from the dynamic system on the network, to the network itself. Since fundamental properties, such as the transitivity of edges, do not necessarily hold in temporal networks, many of these methods need to be quite different from those for static networks (70).

2.5 Conclusions

In this section we recalled the basic notions of graph theory. In particular, we presented the very fundamental tools to analyze and characterize networked systems. We gave a general overview of some data-driven networks to provide to the reader some examples of how a network approach can be used to describe very different settings. In chapter 4 and 5 we will use a network representation to describe the world-wide airport network and the local commuting patterns. Then, we briefly introduced three network models that are the prototype examples of graphs with different statistical behavior: random graphs, small-world networks, scale-free networks. These models are often used as benchmarks to analyze new datasets and to understand the aggregation mechanism of different systems. We finally introduced the concept of dynamical network, that will be crucial in the chapter 6 where we present the longitudinal analysis of a dynamical network and we discuss some important features that can enhance our understanding of the unfolding of dynamical processes on networked systems with temporal dimension.

3

Theoretical Framework: Epidemic Models

Contents

The etiology and the mechanism of transmission of infectious pathogens has been successfully investigated for many diseases. In general, diseases transmitted by viral agents, such as influenza, measles and chicken pox, confer immunity against reinfection, while diseases transmitted by bacteria, such as tuberculosis, meningitis, and gonorrhea, do not led to a lifelong immunity. Other diseases, such as malaria or West Nile virus, are not directly transmitted from hosts to hosts but by vectors, which are agents who carry the infections between the hosts. For sexually transmitted infections the disease is transmitted back and forth between the individuals by means of sex acts. Different diseases call for different theoretical designs and in this chapter we will present the general framework to investigate the unfolding of airborne diseases such as an influenza-like-illness (ILI) by means of a mathematical model. It is worth to notice that the mathematical description of an infectious disease can explore different scales depending on the observables under study. The evolution of the virus population within the host, the number of children or elderly people that would be hospitalized during an epidemic wave or the number of country reached by the epidemic are crucial issues that has to be addressed with different

models. In this thesis we investigate the disease transmission between individuals in a closed population. In chapter 5 we will present the results of our studies on the 2009 H1N1 pandemic influenza where we focused on the country level by considering the world-wide dissemination of the disease. In chapter 6 we assess the impact of an emerging infectious disease among premises of domestic animals.
In this chapter we will introduce the very fundamental concepts of this field but we refer the reader for a deeper analysis of the subject to some classical textbooks (71; 72; 73; 74; 75). Similar techniques can be further generalized for modeling the spread of rumors and the cascade effects of economic crises.

3.1 Compartmental models

The first theoretical approach to the propagation of infectious diseases, namely smallpox, was taken by Daniel Bernoulli in the 18th century. Almost two centuries later, Kermack and McKendrick formalized the concept of *compartmental models* by using a set of ordinary differential equations to describe the unfolding of an epidemic. This approach became extremely popular and powerful and represents the basis of the modern computational epidemiology.
The compartmental model simply assumes that the individuals of a closed population are divided into a discrete set of compartments according to their health status (71; 76) such as susceptible people who can contract the infection S, those who are already infected and can transmit the disease to other people I and those who have already contracted the infection and have recovered from the disease R. In this simple case the model is named SIR, but if the disease does not confer any long lasting immunity the SIS model would be more appropriate, or if the disease under study lead to a permanent infection it can be used the SI model. Additional stages of the disease can be introduced depending on the type of the disease. Examples of these extensions will be shown in details in the chapter 4. With a compartmental model is possible to tackle the key questions of infectious disease spreading:

- if a new pathogen is introduced in a "virgin" population, does this cause an epidemic?
- if so, with what rate does the number of infected hosts increase during the rise of the epidemic?
- what proportion of the population will ultimately have experienced infection?

Let us consider a population of N individuals and let us define the number of individuals in the class $[m]$ at the time t as $X^{[m]}(t)$. Since we are not considering births and deaths, we assume a conservation of the number of individuals

$$N = \sum_m X^{[m]}(t). \tag{3.1}$$

The transitions between different compartments depend on the specific disease that we are modeling. In general, the transition from one compartment to the other is specified by a reaction rate that depends on the disease etiology, such as the infection transmission rate or the recovery rate. In compartmental models there are two possible types of elementary processes ruling the disease dynamics. We can imagine the disease transmission as a reaction process where the rate of interaction of two different subsets of the population is proportional to the product of the numbers in each of the subset concerned, while the spontaneous recovery process occurs with a constant rate.

$$S + I \to 2I \qquad I \to R \tag{3.2}$$

The first process can be generally described by considering the variation of individuals in the class m as $\sum_{h,g} \nu^m_{h,g} a_{h,g} X^{[h]} X^{[g]} N^{-1}$, where $a_{h,g}$ is the transition rate of the process and $\nu^m_{g,h} = [-1, 0, 1]$ the change in the number of $X^{[m]}$ due to the interaction. The factor N^{-1} follow from the homogeneous approximation, where the interaction of each individual of class $[h]$ with individuals of class $[g]$ depends only on the density of individuals of such class $X^{[g]}/N$. The homogeneous approximation is therefore equivalent to the mean-field one used for physical models and considers an effective interaction, a mass-action law, determining the force of infection in the same way for all the individuals of the system. The spontaneous transition of one individual from one compartment $[m]$ to another one $[h]$ is given by $\sum_h \nu^m_h a_h X^{[h]}$ where $\nu^m_h = [-1, 0, 1]$ and a_h is the transition rate. We can now write the general deterministic reaction rate equations for the quantity $X^{[m]}$ summing the two contributions presented:

$$\partial_t X^{[m]} = \sum_{h,g} \nu^m_{h,g} a_{h,g} X^{[h]} X^{[g]} N^{-1} + \sum_h \nu^m_h a_h X^{[h]}. \tag{3.3}$$

Equation 3.3 represents the general expression to define many epidemiological models. In particular it would be easy to derive the differential equations governing the SI, SIS and SIR models. In the following we will focus on the SIR model, since in this work we have always in mind diseases that lead to a permanently recovered state.

Considering the SIR compartmentalization and an infection transmission rate β and a recovery rate μ, the system is thus governed by the following differential equations providing the variations of density of susceptible $s(t) = S(t)/N$, density of infectious $i(t) = I(t)/N$ and density of recovered individuals $R(t)/N$ in time:

$$\begin{aligned} \frac{ds(t)}{dt} &= -\beta i(t)s(t) \\ \frac{di(t)}{dt} &= \beta i(t)s(t) - \mu i(t) \\ \frac{dr(t)}{dt} &= \mu i(t) \end{aligned} \tag{3.4}$$

The transmissibility β introduced here is a mathematical parameter to model in an effective way the rate of infection of susceptible hosts homogeneously mixed with infectious individuals averaging all the biological, social and environmental factors contributing to disseminate or hamper the spread of the virus. For a sake of simplicity, the fundamental quantity k, counting the number of contacts that each individual experienced and representing the number of potential individuals that may transmit or acquire the infection has been dropped by assuming that $k = \langle k \rangle$ for all the individuals and thus rescaling accordingly β. Another implicit assumption of this model is that the time scale of the disease is much smaller than the lifespan of individuals; therefore we do not include in the equations terms accounting for the birth or natural death of individuals.
Let us imagine that a new infected individual is introduced in the closed population N. At the early stage of the epidemic we have $i(0) = 1/N$ and we can consider the number of susceptibles $s(0) \simeq 1$ and $r(0) = 0$. Since the variation of $i(t)$ can be written as $di(t)/dt = \mu i(t) + \beta i(t)[1 - r(t) - i(t)]$, we can thus use a linear approximation neglecting all the i^2 terms. The density of infected individuals is:

$$i(t) \simeq i_0 e^{t/\tau} \tag{3.5}$$

where i_0 is the initial density of infected individuals and τ is the typical outbreak time $\tau^{-1} = \beta - \mu$ (71). This expression suggests an important consideration: if the recovery rate is greater than the transmission rate, τ assumes negative values and the number of infected individuals fade out on the timescale $|\tau|$. The ratio β/μ is thus a crucial quantity and represents the *epidemic threshold*. Defining the *basic reproductive number* R_0, as

$$R_0 = \frac{\beta}{\mu}, \tag{3.6}$$

an infectious disease is able to spread in a large part of the population only if $R_0 > 1$ i.e. when the exponent of equation 3.6 greater than zero. The basic reproductive number represents the

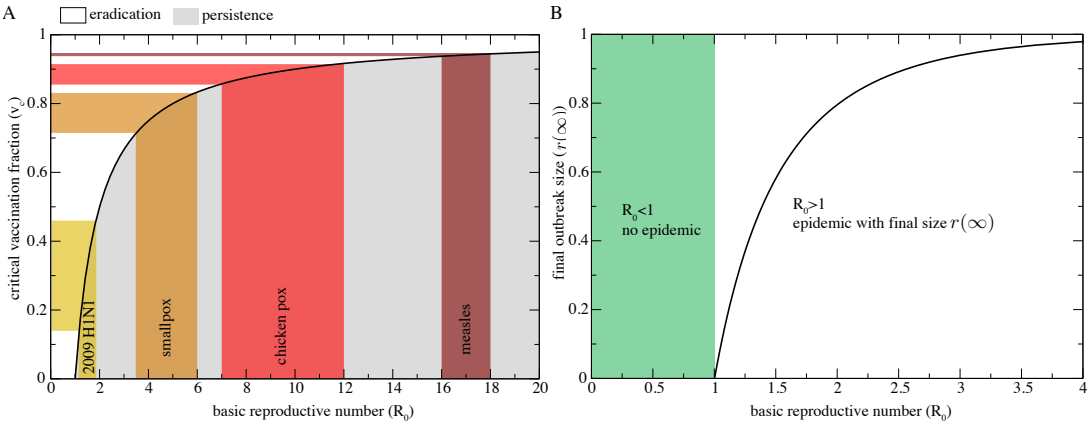

Figure 3.1: (A) Eradication Criterion. The critical fraction of people that has to be vaccinated to eradicate the disease as a function of R_0 is shown. (B) Final outbreak size as a function of the basic reproductive number.

number of new secondary cases that an infected individual may infect in a fully susceptible population before getting recovered.

The relevance of the epidemic threshold is also related to the protection of populations by means of immunization programs. These correspond to vaccination policies aimed at the eradication of the epidemics. Let us imagine that a fraction $v = V/N$ of the original population is vaccinated and thus fully immune to the disease. The initially susceptible individuals are $s(0) = 1 - v - i(0)$, and the linear approximation is therefore:

$$di(t)/dt = [\mu + \beta(1 - v)]i(t) \tag{3.7}$$

The new condition for observing an epidemic outbreak is $R_0' = \frac{\beta}{\mu}(1 - v) > 1$. It is thus possible to define the *critical vaccination fraction*

$$v_c = 1 - \frac{\mu}{\beta} = 1 - \frac{1}{R_0} \tag{3.8}$$

representing the fraction of individuals to immunize to eradicate an infection. As shown in Figure 3.1A, because of a *herd immunity* it is not necessary to vaccinate everyone to eradicate an infection. If the vaccine is not perfect and has an efficacy $0 < E \leq 1$ of preventing the transmission from an infected contact, then $v_c = \frac{1}{E}(1 - \frac{1}{R_0})$. It may happen that for highly infectious diseases (large R_0) and not so effective vaccines (small E), the critical vaccination fraction exceeds 1. This means that vaccination alone cannot prevent an outbreak.

Considering the SIR compartmentalization it is also possible to approximately evaluate the final size of the epidemic (73). Considering the equations 3.4, let us divide the variation of susceptible individuals by the variation of recovered ones:

$$\frac{ds}{dr} = \frac{-\beta s(t) i(t)}{\mu i(t)} = -R_0 s(t) \tag{3.9}$$

Integrating with respect to dr, we obtain $s(t) = s(0)e^{-R_0 r(t)}$. Now, when the epidemic is over, by definition, we have $i(\infty) = 0$ and $s(\infty) = 1 - r(\infty)$. We thus end up with the transcendental equation:

$$r(\infty) = 1 - s(0)e^{-R_0 r(\infty)} \tag{3.10}$$

Ideally, it is possible to estimate the basic reproductive number of a disease using the total number of infected cases using the formula 3.10. In Figure 3.1B the final size is shown as a function of R_0.

It is important to stress that the deterministic continuous equations presented here are valid only in case of sufficiently large populations and in general a more realistic approach able to capture the natural chance effects of the epidemic transmission would require a full stochastic description. To account for this variability the stochastic dynamics rely on an integer-based population and events occur at probabilistic rates. The linear analysis of the SIR model highlights the presence of three basic stages in an epidemic evolution. Initially when few infected individuals are introduced in the population we define a pre-outbreak stage in which the evolution is noisy and dominated by stochastic effects that are extremely relevant in the presence of few contagious events. This is a stage in which epidemics may or may not disappear from the population just because of stochastic effects. When the infected individuals are enough to make stochastic effects negligible, but still very few compared with the whole population, we observe an exponential take off of the infected cases as described by the equation 3.5. Finally, the decrease of susceptible individuals reduces the force of infection of each infected individual and the exponential growth cannot be sustained any longer in the population, we observe the epidemic turn over and the outbreak will ultimately disappear. However, it is worth stressing that while the outbreak will occur with finite probability if the parameters poise the system above the epidemic threshold, this probability is not equal to one. Actually the stochastic fluctuations may lead to the extinction of the epidemics even well above the epidemic threshold. It has been shown (72) that the extinction probability of an epidemic starting with I_0 infected individuals is equal to $R_0^{-I_0}$. For instance, in the case of a single infected individual, even for values of R_0 as

high as 2 the outbreak probability is just 50%.

3.2 Epidemic spreading on graphs

In the previous section we have presented the simple compartmental epidemic models to investigate an infectious disease spreading in a homogeneous population disregarding every possible substructure (gender, age, risk groups, etc...). As already mention, the parameter β governs in an effective way the transmission rate of the disease between infectious and susceptible individuals. In particular, the approximation adopted so far where the β value has been rescaled in order to disregard the usually unknown average number of contacts $\langle k \rangle$ through which the disease can be transmitted, is equivalent to consider the spreading on a random graph. Nevertheless, many real social and technological networks of epidemiological relevance (mobility networks, the web of sexual contacts, internet, etc...) are far for being homogeneous and the hypothesis that each individual in the system has the same number of connections $k \simeq \langle k \rangle$ might not be a good approximation.

In this section, by describing the epidemic spreading on a networked system, we show that the fluctuations play a main role in determining the epidemic properties and the spreading may be favored in heterogeneous networks (27; 71; 77; 78). Let us consider an uncorrelated graph completely defined by the degree distribution $P(k)$, and let us divide the nodes according to their health status. The disease can be transmitted from one node to the other only if the nodes are connected through an edge. In order to take into account the heterogeneity induced by the presence of nodes with different connectivity, here we will use a degree block approximation (77; 79; 80): all nodes with the same degree are statistically equivalent. Thus our results will not apply to structured networks in which a distance or a time ordering can be defined; for instance, when the small-world property is not present (79; 81). Here, we have to relax the homogeneous mixing hypothesis made in the previous section and leading to equation 3.4 and work instead with the relative density of infected and susceptible vertices with given degree k:

$$i_k = \frac{I_k}{N_k}, \quad s_k = \frac{S_k}{N_k}. \tag{3.11}$$

and the global averages are given by:

$$i = \sum_k P(k) i_k \quad , s = \sum_k P(k) s_k. \tag{3.12}$$

Let us first consider the very simple SI model. In this case we know that the whole connected component of the system will be infected independently of the spreading rate, but it is very interesting to see the effect of topological fluctuations on the spreading velocity. Considering the class of degree k and defining $\theta_k(t)$ the density of infected neighbors of vertices of degree k the evolution equations read:

$$d_t i_k(t) = \beta[1 - i_k(t)]k\theta_k(t). \tag{3.13}$$

The term θ_k represents the average probability that any given neighbor of a node of degree k is infected, in the homogeneous assumption it is equal to the density of infected nodes. By considering that at least one of the edges of each infected vertex points to another infected vertex from which the infection has been transmitted, the most general expression for θ_k is

$$\theta_k = \sum_{k'} \frac{k'-1}{k'} P(k'|k) i_{k'}. \tag{3.14}$$

The simplest case we can analyze is a network with no degree correlations meaning that the probability that an edge departing from a vertex of degree k arrives at a vertex of degree k' is independent from the degree of the initial vertex k. In this case the conditional probability does not depend on the originating node and it is possible to show that $P(k'|k) = k'P(k')/\langle k\rangle$ and thus

$$\theta_k(t) = \theta(t) = \frac{\sum_{k'}(k'-1)P(k')i_{k'}(t)}{\langle k\rangle}. \tag{3.15}$$

Using this in 3.13 and neglecting the terms i^2, we have:

$$d_t i_k(t) = \beta k \theta(t), \tag{3.16}$$

multiplying both sides of this expression by $(k-1)P(k)$ and summing over k we get:

$$d_t \theta(t) = \beta\theta(t)\left(\frac{\langle k^2\rangle}{\langle k\rangle} - 1\right). \tag{3.17}$$

We can solve these equations fixing $i_k(t=0) = i_0$ getting:

$$i_k(t) = i_0\left[1 + \frac{k(\langle k\rangle - 1)}{\langle k^2\rangle - \langle k\rangle}(e^{t/\tau} - 1)\right], \tag{3.18}$$

with

$$\tau = \frac{\langle k\rangle}{\beta(\langle k^2\rangle - \langle k\rangle)}. \tag{3.19}$$

It is clear that the fraction of infected individuals increases exponentially. This process is faster for high degree nodes. The growth time scale is measured by the heterogeneity ratio $\langle k^2\rangle/\langle k\rangle$. For scale free networks with power law degree distributions $P(k) = k^{-\gamma}$ with exponent $2 < \gamma \leq 3$

in the limit $N \to \infty$ we have an unbounded second moment, then in uncorrelated scale-free networks we would have a virtually instantaneous rise of the epidemic size. The reason for that is quite intuitive. Once the disease has reached the hubs it can spread rapidly among the network. Multiplying both sides by $P(k)$ and summing over all k we get:

$$i(t) = i_0 \left[1 + \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} (e^{t/\tau} - 1) \right]. \tag{3.20}$$

The above results can be easily extended to the SIR model. Here the variation of infected individuals of degree k has to take into account also the recovery process:

$$d_t i_k(t) = \beta k s_k(t) \theta_k(t) - \mu i_k(t), \tag{3.21}$$

where $s_k(t) = 1 - r_k(t) - i_k(t)$. Again considering the linear approximation and uncorrelated networks we get the time scale τ:

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - (\mu + \beta) \langle k \rangle}. \tag{3.22}$$

The fluctuations are again very important, and here they play a crucial role in the definition of the epidemic threshold. In order to ensure an epidemic outbreak the condition $\tau > 0$ must be satisfied:

$$\frac{\beta}{\mu} \geq \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{3.23}$$

For scale-free networks with exponent $2 < \gamma \leq 3$ in the limit of infinite size the second moment diverges, so we have a null epidemic threshold. This is an important result that confirms how heterogeneous networks behave in a completely different way from homogeneous networks. Scale-free networks are then an ideal topology for the spreading of infectious diseases.

3.3 Metapopulation models

In the previous section we studied systems with a homogeneous mixing approximation or structured populations in which each node of the network corresponds to a single individual. Recently the effect of heterogeneous connectivity patterns has been studied in the case in which each node of the system may be occupied by any number of particles and the connections allow for the displacement of particles form one node to the other (82). In an epidemic framework, particles represent hosts moving between different locations, such as cities or urban areas called, in general, subpopulations. These models are called metapopulations epidemic models and can be formalized on different theoretical substructures, from regular lattices to random graphs, but in the last

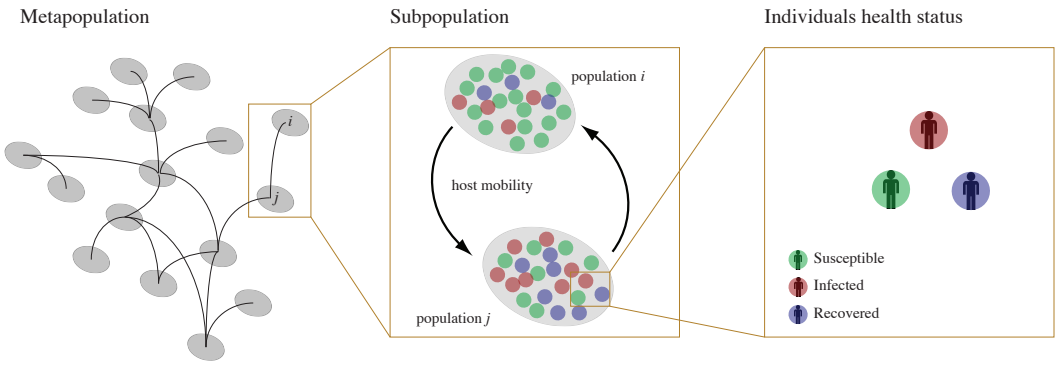

Figure 3.2: Representation of a metapopulation model. Each node of the system contains a population of individuals who are characterized with respect to their stage of the disease. In this case we are considering Susceptible, Infected and Recovered indicated in different colors in the picture. Individuals can diffuse from a subpopulation/node to another on the network of connections among subpopulations.

years the abundance of data-driven networks which trace the activities of individuals have led to models based on the detailed knowledge of the spatial structure of the environment and of transportation infrastructures, movement patterns and traffic networks (78; 83; 84; 85; 86; 87; 88). In this framework, the nodes of the network are the subpopulations and the coupling among them is shaped by the connectivity patterns represented by the network topology resulting by the movement of individuals from one subpopulation to the other. A sketch of the metapopulation approach is shown in Figure 3.2. Each node i is connected to other k_i nodes according to its degree resulting in a network with degree distribution $P(k)$ and distribution moments $\langle k^\alpha \rangle = \sum_k k^\alpha P(k)$.

Realistic descriptions are provided by explicit mechanistic approaches, in which detailed rates of traveling/commuting obtained from data, or from empirical fit to gravity law models, are included (78; 89). A typical assumption is to consider the diffusion process as Markovian implying that the movements of individuals have no memory. Individuals are not labeled according to their original subpopulation, so they move without having memory of their origin. At each time step the movement of individuals is given according to a matrix p_{ij} that encodes the probability that an individual in the subpopulation i will travel to the subpopulation j. Being w_{ij} the traffic among subpopulations and N_i the number of individuals living in the node i, we define

$$p_{ij} \sim \frac{w_{ij}}{N_i}. \tag{3.24}$$

These probabilities in realistic models are obtained from real data (45; 78; 90; 91; 92; 93; 94; 95; 96).

Individuals in the same location may get in contact and interact according to the infection dynamics modeled as a reaction process. Within each subpopulation they are divided into classes denoting their health status according to the modeled disease (71). A key point is to evaluate the force of infection generated by the infectious individuals in subpopulation j on the individuals in subpopulation i (84; 85; 97; 98; 99; 100). In the case of a simple SIR model for the evolution of the disease, the metapopulation approach amounts to writing, for each subpopulation, equations such as:

$$\Delta I_i(t) = f(I_i, S_i, R_i) + \Omega_i(I) \tag{3.25}$$

where the first term represents the variation of infected individuals due to the infection dynamics within the subpopulation i and the second term corresponds to the net balance of infectious individuals traveling in and out of the city i. This last term, the transport operator Ω_i, depends on the probability p_{ij} that an infected individual will go from city i to city j and can be generally written as:

$$\Omega_i(I) = \sum_i (p_{ji} I_j - p_{ij} I_i) \tag{3.26}$$

representing the total sum of infectious individuals arriving in subpopulation i from all connected subpopulations j, minus the amount of individuals traveling in the opposite directions. Similar equations can be written for all the compartments included in the disease model, finally leading to a set of equations where the transport operator acts as a coupling term among the evolution of the epidemics in the various subpopulations.

3.4 Conclusion

We are aware that the transmission of airborne diseases is influenced by many biological, social and environmental factors. For instance, the infectivity of a diseased individual depends on the viral load (101; 102) as well as the susceptibility is influenced by the individual antibody response. The age-specific contact patterns (103; 104) have been found to be relevant for infections transmitted by the respiratory or close-contact route and, finally, it has been recognized that the absolute air humidity may contribute to disseminate viruses in aerosolized droplets (105; 106; 107). Nevertheless, with the basic compartmental description introduced in this chapter, where all these factors are disregarded or flattened out, it is possible to achieve important insights about an infectious disease spreading. Such approach represent the first step in the epidemic modeling,

and we discussed how through the homogeneous mixing approximation it is possible to derive a mathematical description of the epidemic threshold, the final size of an epidemic outbreak and the critical vaccination fraction. We then briefly described how the modeling of disease spreading on networked settings can take advantage of the degree-block approximation in order to take into account the heterogeneity of the system due to different contact patterns. This approach will be crucial in chapter 5 for the analytical assessment of the disease containment by means of human mobility restrictions. Furthermore, in chapter 6, leveraging on a network description, we will present a systematic investigation of an emerging infectious disease spreading through the Italian livestock premises. Finally, the metapopulation framework introduced in the last section represents the starting point for presenting the Global Epidemic and Mobility (GLEaM) model described in chapter 4 and chapter 5.

4

GLobal Epidemic and Mobility model

Contents

In this chapter we present in detail the Global Epidemic and Mobility model (GLEaM), that is a discrete stochastic epidemic computational model based on a meta-population approach able to perform numerical simulations of global epidemic spreading. The design and implementation of GLEaM started in 2005 and in the last years it involved almost twenty collaborators from different laboratories in Europe and the US. Along with the academic research, the GLEaMviz project (`www.gleamviz.org`) is devoted to provide a public software with a user friendly interface for the simulation of large-scale epidemic outbreaks and to participate to outreach activities about emerging infectious diseases spreading. During my PhD training, within the Computational Epidemiology Laboratory at the I.S.I. Foundation, Turin, IT, I have contributed to the development of GLEaM by collecting and analyzing part of the short range mobility data in-

cluded in the model, and by testing and implementing the computational infrastructure described in this chapter. GLEaM is coded in C/C++ and runs conveniently on high-end desktop machines. The results presented in chapter 5 were achieved by using GLEaM to investigate the H1N1 2009 influenza pandemic. In this context, I have performed part of the simulations for the estimation of the disease parameters as described in 5.2 and I have performed the numerical estimate of the number of cases in Mexico in the early phase of the outbreak as discussed in 5.4. I have implemented the code, performed the simulations and analyzed the results to test different intervention strategies as described in sections 5.5.1 and 5.5.3 and I have helped in the analysis and the assessment of the critical care demand presented in 5.5.2. Furthermore I have performed the simulations and provided the data for the implementation of the "Epidemic Planet" scientific exhibit (`http://www.gleamviz.org/outreach-activities/`) that was hosted at the International Science Festival that took place in Edinburgh, UK, April 3 - 17, 2010 and at the International Conference for High Performance Computing, Networking, Storage and Analysis that took place in New Orleans, LA, USA, November 13 - 19, 2010

Structured metapopulation model

Here we present the detailed definition and data description of the global structured metapopulation model. The computational model is based on three data/model layers. The first layer is a data layer defining the census area and the subpopulation structure. The second one refers to human mobility model defined by the transportation and commuting networks characterizing the interactions and exchanges of individuals across subpopulations. The third layer is the epidemic dynamic model that defines the evolution of the infectious disease inside each subpopulations.

4.1 Global Population and subpopulations definition

The population dataset was obtained from the Web sites of the "Gridded Population of the World" and the "Global Urban-Rural Mapping" projects (108; 109), which are run by the Socioeconomic Data and Application Center (SEDAC) of Columbia University. The surface of the world is divided into a grid of cells that can have different resolution levels. Each of these cells has assigned an estimated population value.

Out of the possible resolutions, we have opted for cells of 15×15 minutes of arc to constitute the basis of our model. This corresponds to an area of each cell approximately equivalent to a

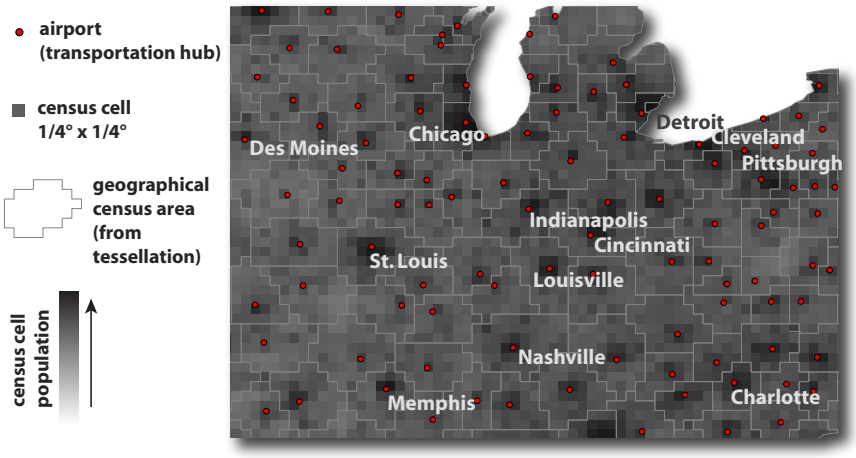

Figure 4.1: Population database and Voronoi tessellation around main tranportation hubs. The world surface is represented in a grid-like partition where each cell - corresponding to a population values - is assigned to the closest airport. Geographical census areas emerge that constitute the sub-populations of the meta-population model.

rectangle of $25 \times 25\, kms$ along the Equator. The dataset comprises 823 680 cells, of which 250 206 are populated. Since the coordinates of each cell center and those of the airports are known, the distance between the cells and the airports can be calculated. We have performed a Voronoi-like tessellation of the Earth surface assigning each cell to the closest airport that satisfies the following two conditions: (i) Each cell is assigned to the closest airport within the same country, and (ii), the distance between the airport and the cell cannot be longer than $200\, kms$. This cutoff naturally emerges from the distribution of distances between cells and closest airports, and it is introduced to avoid that in barely populated areas such as Siberia we can generate geographical census areas thousands of kilometer wide but with almost no population. It also corresponds to a reasonable upper cutoff for the ground traveling distance expected to be covered to reach an airport before traveling by plane.

Before proceeding with the tessellation, we need to take into account that some urban areas include more than one airport. For instance, London has up to six airports, Paris has two, and New York City has three. Our aim is to build a metapopulation model whose subpopulations correspond to the geographical census areas obtained from tessellation. Inside these geographical census areas a homogeneous mixing is assumed. The groups of airports that serve the same urban area need therefore to be aggregated since the mixing within the given urban area is expected to

be high and cannot be represented in terms of separated subpopulations for each of the airports serving the same city. We have searched for groups of airports located close to each other and we manually processed the identified groups of airports to select those belonging to the same urban area. The airports of the same group are then aggregated in a single "super-hub". An example with the final result of the Voronoi tessellation procedure with cells and airports can be seen in Figure 4.1. The geographical census areas become thus the basic subpopulations of our metapopulation model. Their connections will determine the geographical spreading of an hypothetical epidemic.

4.2 World Airport Network

The subpopulations are connected to each other by human mobility fluxes. The long range connections due to the air travels are incorporated in the model by considering the World-wide Airport Network (WAN). The WAN is composed of 3362 commercial airports indexed by the International Air Transport Association (IATA) that are located in 220 different countries. The database contains the number of available seats per year for each direct connection between two of these airports. The coverage of the dataset is estimated to be 99% of the global commercial traffic. The WAN can be seen as a weighted graph comprising 16 846 edges where the weight $\omega_{j\ell}$ of a link (j, ℓ) represents the passenger flow between airports j and ℓ. The network shows a high degree of heterogeneity both in the number of destinations per airport and in the number of passengers per connection (15; 88; 110; 111).

4.3 Commuting Networks

The commuting databases have been collected from the Offices of Statistics of 28 countries in the 5 populated continents. The full dataset comprehends more than 78 000 administrative regions and over five million commuting flow connections between them (see (78))4.1. The definition of administrative unit and the granularity level at which the commuting data are provided enormously vary from country to country. For example, most European countries adhere to a practice that ranks administrative divisions in terms of geocoding for statistical purposes, the so called Nomenclature of Territorial Units for Statistics (NUTS). Most countries in the European Union are partitioned into three NUTS levels which usually range from states to provinces. The commuting data at this level of resolution is therefore strongly coarse-grained. In order to have a higher geographical resolution of the commuting datasets that could match

Table 4.1: Commuting networks in each continent. Number of countries (N_c), number of administrative units (V) and inter-links between them (E) are summarized.

Continent	N_c	V	E
Europe	17	65880	4490650
North America	2	6986	182255
Latin America	4	1858	63678
Asia	3	2732	323815
Oceania	2	746	30679
Total	28	78202	5091077

the resolution scale of our geographical census areas, we looked for smaller local administrative units (LAU) in Europe. The US or Canada report commuting at the level of counties. However, even within a single country the actual extension, shape, and population of the administrative divisions are usually a consequence of historical reasons and can be strongly heterogeneous.

Such heterogeneity renders the efforts to define a universal law describing commuting flows likely to fail. The mobility behavior might indeed result different across countries simply due to the country specific partition of the population into administrative boundaries. In order to overcome this problem, and in particular to define a data-driven short range commuting for GLEaM, we used the geographical census areas obtained from the Voronoi tessellation as the elementary units to define the centers of gravity for the process of commuting. This allows to deal with similar units across the world wth respect to mobility as emerged from a tessellation around main hubs of mobility and not country specific administrative boundaries. We have therefore mapped the different levels of commuting data into the geographical census areas formed by the Voronoi-like tessellation procedure described above. The mapped commuting flows can be seen as a second transport network connecting subpopulations that are geographically close. This second network can be overlaid to the WAN in a multi-scale fashion to simulate realistic scenarios for disease spreading. A schematic illustration of the different layers of the model is shown in figure 4.2. The network exhibits important variability in the number of commuters on each connection as well as in the total number of commuters per geographical census area. Since the census areas are relatively homogeneous and similar we can estimate a gravity law that successfully reproduces the commuting data obtained across different continents, and provide us

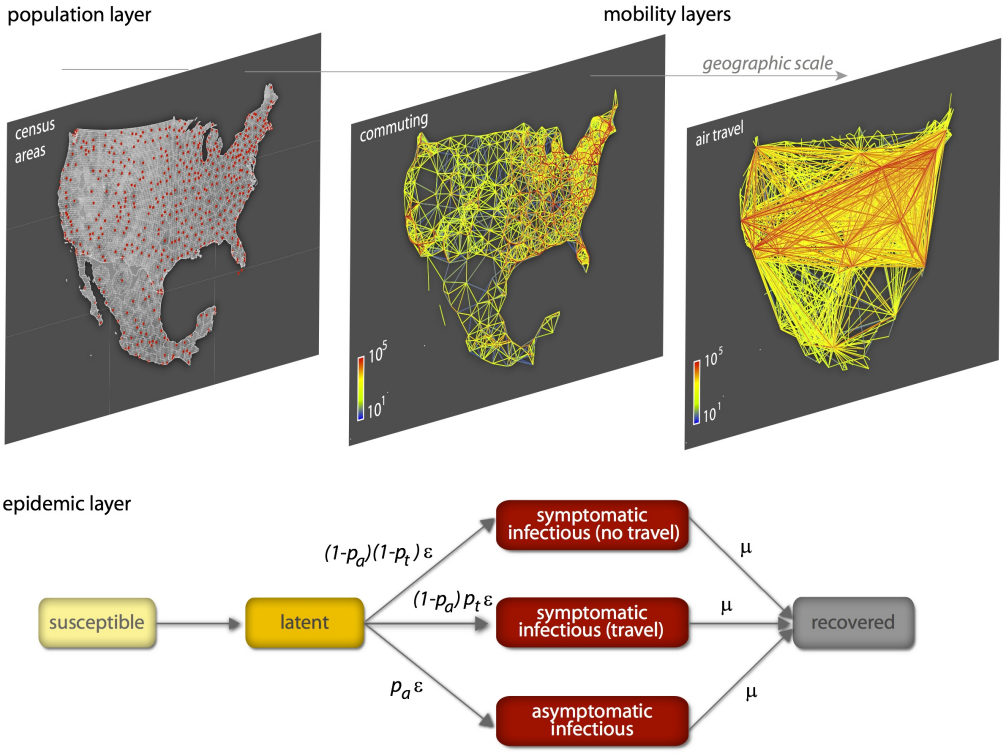

Figure 4.2: Schematic illustration of the GLobal Epidemic and Mobility (GLEaM) model. Top: census and mobility layers that define the subpopulations and the various types of mobility among those (commuting patterns and air travel flows). The same resolution is used worldwide. Bottom: compartmental structure in each subpopulation. A susceptible individual in contact with a symptomatic or asymptomatic infectious person contracts the infection at rate β or $r_\beta\beta$, respectively, and enters the latent compartment where he is infected but not yet infectious. At the end of the latency period, each latent individual becomes infectious, entering the symptomatic compartments with probability $1-p_a$ or becoming asymptomatic with probability p_a. The symptomatic cases are further divided between those who are allowed to travel (with probability p_t) and those who would stop traveling when ill (with probability $1-p_t$). Infectious individuals recover permanently with rate μ. All transition processes are modeled through multinomial processes.

with estimations for the possible commuting levels in the countries for which such data is not available as in ref. (78).

4.4 Epidemic model

Each geographical census area corresponds to a subpopulation in the metapopulation model, inside which we consider a Susceptible-Latent-Infectious-Recovered (SLIR) compartmental scheme, typical of influenza-like illnesses (ILIs), where each individual has a discrete disease state assigned at each moment in time. In Fig. 4.2, a diagram of the compartmental structure with transitions between compartments is shown. The contagion process, i.e. generation of new infections, is the only transition mechanism which is altered by short-range mobility, whereas all the other transitions between compartments are spontaneous and remain unaffected by the commuting. The rate at which a susceptible individual in subpopulation j acquires the infection, the so called force of infection λ_j, is determined by interactions with infectious persons either in the home subpopulation j or in its neighboring subpopulations on the commuting network.

Given the force of infection λ_j in subpopulation j, each person in the susceptible compartment (S_j) contracts the infection with probability $\lambda_j \Delta t$ and enters the latent compartment (L_j), where Δt is the time interval considered. Latent individuals exit the compartment with probability $\varepsilon \Delta t$, and transit to asymptomatic infectious compartment (I_j^a) with probability p_a or, with the complementary probability $1 - p_a$, become symptomatic infectious. Infectious persons with symptoms are further divided between those who can travel (I_j^t), with probability p_t, and those who are travel-restricted (I_j^{nt}) with probability $1 - p_t$. All the infectious persons permanently recover with probability $\mu \Delta t$, entering the recovered compartment (R_j) in the next time step. All transitions and corresponding rates are summarized in Table 4.2 and in Figure 4.2. In each subpopulation the variation of the number of individuals in each compartment $[m]$ can be written at any given time step as

$$X_j^{[m]}(t + \Delta t) - X_j^{[m]}(t) = \Delta X_j^{[m]} + \Omega_j([m]) \tag{4.1}$$

where the term $\Delta X_j^{[m]}$ represents the change due to the compartment transitions induced by the disease dynamics and the transport operator $\Omega_j([m])$ represents the variations due to the traveling and mobility of individuals. The latter operator takes into account the long-range airline mobility and defines the minimal time scale of integration to 1 day. The mobility due to the commuting flows is taken into account by defining effective force of infections by using a time scale separation approximation as detailed in the following sections.

Table 4.2: Transitions between compartments and their rates.

Transition	Type	Rate
$S_j \to L_j$	Contagion	λ_j
$L_j \to I_j^a$	Spontaneous	εp_a
$L_j \to I_j^t$	"	$\varepsilon(1-p_a)p_t$
$L_j \to I_j^{nt}$	"	$\varepsilon(1-p_a)(1-p_t)$
$I_j^a \to R_j$	"	μ
$I_j^t \to R_j$	"	μ
$I_j^{nt} \to R_j$	"	μ

4.5 Stochastic and discrete integration of the disease dynamics

In each subpopulation j, we define an operator acting on a compartment $[m]$ to account for all the transitions out of the compartment in the time interval Δt. Each element $\mathcal{D}_j([m],[n])$ of this operator is a random variable extracted from a multinomial distribution and determines the number of transitions from compartment $[m]$ to $[n]$ occurring in Δt. The change $\Delta X_j^{[m]}$ of a compartment $[m]$ in this time interval is given by a sum over all random variables $\{\mathcal{D}_j([m],[n])\}$ as follows

$$\Delta X_j^{[m]} = \sum_{[n]} \{-\mathcal{D}_j([m],[n]) + \mathcal{D}_j([n],[m])\} \quad . \tag{4.2}$$

As a concrete example let us consider the evolution of the latent compartment. There are three possible transitions from the compartment: transitions to the asymptomatic infectious, the symptomatic traveling and the non-traveling infectious compartments. The elements of the operator acting on L_j are extracted from the multinomial distribution

$$Pr^{Multin}(L_j(t), p_{L_j \to I_j^a}, p_{L_j \to I_j^t}, p_{L \to I_j^{nt}}) \tag{4.3}$$

determined by the transition probabilities

$$\begin{aligned} p_{L_j \to I_j^a} &= \varepsilon p_a \Delta t \quad , \\ p_{L_j \to I_j^t} &= \varepsilon(1-p_a)p_t \Delta t \quad , \\ p_{L \to I_j^{nt}} &= \varepsilon(1-p_a)(1-p_t)\Delta t \quad , \end{aligned} \tag{4.4}$$

and by the number of individuals in the compartment $L_j(t)$ (its size). All these transitions cause a reduction in the size of the compartment. The increase in the compartment population is due to the transitions from susceptibles into latents. This is also a random number extracted from a binomial distribution

$$Pr^{Bin}(S_j(t), p_{S_j \to L_j}) \tag{4.5}$$

given by the chance of contagion

$$p_{S_j \to L_j} \quad = \quad \lambda_j \Delta t \quad , \tag{4.6}$$

with a number of attempts given by the number of susceptibles $S_j(t)$. After extracting these numbers from the appropriate multinomial distributions, we can calculate the change $\Delta L_j(t)$ as

$$\Delta L_j(t) = L_j(t+1) - L_j(t) = -\left[\mathcal{D}_j(L, I^a) + \mathcal{D}_j(L, I^t) + \mathcal{D}_j(L, I^{nt})\right] + \mathcal{D}_j(S, L) \; . \tag{4.7}$$

4.6 The integration of the transport operator

The transport operator is defined by the airline transportation data and sets the integration time scale to 1 day. The number of individuals in the compartment $[m]$ traveling from the subpopulation j to the subpopulation ℓ is an integer random variable, in that each of the X_j potential travelers has a probability $p_{j\ell} = w_{j\ell}/N_j$ to go from j to ℓ. In each subpopulation j the numbers of individuals $\xi_{j\ell}$ traveling on each connection $j \to \ell$ at time t define a set of stochastic variables which follows the multinomial distribution

$$P(\{\xi_{j\ell}\}) = \frac{X_j^{[m]}!}{(X_j^{[m]} - \sum_\ell \xi_{j\ell})! \prod_\ell \xi_{j\ell}!} (1 - \sum_\ell p_{j\ell})^{(X_j^{[m]} - \sum_\ell \xi_{j\ell})} \prod_\ell p_{j\ell}^{\xi_{j\ell}}, \tag{4.8}$$

where $(1 - \sum_\ell p_{j\ell})$ is the probability of not traveling, and $(X_j^{[m]} - \sum_\ell \xi_{j\ell})$ identifies the number of non traveling individuals of the compartment $[m]$. We use standard numerical subroutines to generate random numbers of travelers following these distributions. The transport operator in each subpopulation j is therefore written as

$$\Omega_j([m]) = \sum_\ell (\xi_{\ell j}(X_\ell^{[m]}) - \xi_{j\ell}(X_j^{[m]})), \tag{4.9}$$

where the mean and variance of the stochastic variables are $\langle \xi_{j\ell}(X_j^{[m]}) \rangle = p_{j\ell} X_j^{[m]}$ and $\mathrm{Var}(\xi_{j\ell}(X_j^{[m]})) = p_{j\ell}(1 - p_{j\ell}) X_j^{[m]}$. Direct flights as well as connecting flights up to two-legs flights can be considered. It is worth remarking that on average the airline network flows are balanced so that the subpopulations N_j are constant in time, e.g. $\sum_{[m]} \Omega_j([m]) = 0$.

4.7 Time-scale separation and the integration of the commuting flows

The Global Epidemic and Mobility (GLEaM) modeler combines the infection dynamics with long- and short-range human mobility. Each of these dynamical processes operates at a different time scale. For ILI there are two important intrinsic time scales, given by the latency period ε^{-1} and the duration of infectiousness μ^{-1}, both larger than 1 *day*. The long-range mobility given by the airline network has a time scale of the order of 1 *day*, while the commuting takes place in a time scale of approximately $\tau^{-1} \sim 1/3$ *day*. The explicit implementation of the commuting in the model thus requires a time interval shorter than the minimal time of airline transportation. In particular, for the numerical simulations we would use a time step of the order of 1/3 *day* or less and we should explicitly consider the displacements and interactions of individuals in the different epidemic compartments that spend a fraction of day in a close location. This would triple the number of time steps for each simulation and would enormously increase the complexity of the model design, since the population in each location i should be divided in $[m]$ compartments and further tracked to k close locations to perform the back-and-forth daily commuting. We would consider $[m] \cdot (k+1)$ compartments depending both on the $[m]$ health status and the k commuting route performed daily (the term $+1$ accounts for people that do not commute). Finally, we should aggregate again all the information to investigate the topic under study that is typically longer than 1/3 *day*. To overcome these complications, we use a time-scale separation technique, in which the short-time dynamics is integrated into an effective force of infection in each subpopulation. We start by considering the temporal evolution of subpopulations linked only by commuting flows and evaluate the relaxation time to an equilibrium configuration. Consider the subpopulation j coupled by commuting to other n subpopulations. The commuting rate between the subpopulation j and each of its neighbors i will be given by σ_{ji}. The return rate of commuting individuals is set to be τ. Following the work of Sattenspiel and Dietz (112), we can divide the individuals originary from the subpopulation j, N_j, between $N_{jj}(t)$ who are from j and located in j at time t and those, $N_{ji}(t)$, that are from j and located in a neighboring subpopulation i at time t. Note that by consistency

$$N_j = N_{jj}(t) + \sum_i N_{ji}(t). \tag{4.10}$$

The rate equations for the subpopulation size evolution are then

$$\begin{aligned} \partial_t N_{jj} = & \quad -\textstyle\sum_i \sigma_{ji} N_{jj}(t) + \tau \sum_i N_{ji}(t) \quad , \\ \partial_t N_{ji} = & \quad \sigma_{ji} N_{jj}(t) - \tau N_{ji}(t) \quad . \end{aligned} \tag{4.11}$$

By using condition (4.10), we can derive the closed expression

$$\partial_t N_{jj} + (\tau + \sigma_j)\, N_{jj}(t) = N_j \tau \quad , \tag{4.12}$$

where σ_j denotes the total commuting rate of population j, $\sigma_j = \sum_i \sigma_{ji}$. $N_{jj}(t)$ can be expressed as

$$N_{jj}(t) = e^{-(\tau+\sigma_j)t} \left(C_{jj} + N_j \tau \int_0^t e^{(\tau+\sigma_j)s} ds \right) \quad , \tag{4.13}$$

where the constant C_{jj} is determined from the initial conditions, $N_{jj}(0)$. The solution for $N_{jj}(t)$ is then

$$N_{jj}(t) = \frac{N_j}{(1+\sigma_j/\tau)} + \left(N_{jj}(0) - \frac{N_j}{(1+\sigma_j/\tau)} \right) e^{-\tau(1+\sigma_j/\tau)t} \quad . \tag{4.14}$$

We can similarly solve the differential equation for the time evolution of $N_{ji}(t)$

$$\begin{aligned} N_{ji}(t) \quad = \quad & \frac{N_j \sigma_{ji}/\tau}{(1+\sigma_j/\tau)} - \frac{\sigma_{ij}}{\sigma_j} \left(N_{jj}(0) - \frac{N_j}{(1+\sigma_j/\tau)} \right) e^{-\tau(1+\sigma_j/\tau)t} \\ + \quad & \left[N_{ji}(0) - \frac{N_j \sigma_{ji}/\tau}{(1+\sigma_j/\tau)} + \frac{\sigma_{ij}}{\sigma_j} \left(N_{jj}(0) - \frac{N_j}{(1+\sigma_j/\tau)} \right) \right] e^{-\tau t} \quad . \end{aligned} \tag{4.15}$$

The relaxation to equilibrium of N_{jj} and N_{ji} is thus controlled by the characteristic time $[\tau\,(1+\sigma_j/\tau)]^{-1}$ in the exponentials. Such term is dominated by $1/\tau$ if the relation $\tau \gg \sigma_j$ holds. In our case, $\sigma_j = \sum_i \omega_{ji}/N_j$, that equals the daily total rate of commuting for the population j. Such rate is always smaller than one since only a fraction of the local population is commuting, and it is typically much smaller than $\tau \simeq 3 - 10\, day^{-1}$. Therefore the relaxation characteristic time can be approximated by $1/\tau$. This time is considerably smaller than the typical time for the air connections of one day and hence the approximation of considering the subpopulations $N_{jj}(t)$ and $N_{ji}(t)$ as relaxed to their equilibrium values,

$$N_{jj} = \frac{N_j}{1+\sigma_j/\tau} \quad \text{and} \quad N_{ji} = \frac{N_j \sigma_{ji}/\tau}{1+\sigma_j/\tau} \quad , \tag{4.16}$$

is reasonable. This approximation, originally introduced by Keeling and Rohani (113), allows us to consider each subpopulation j as having an effective number of individuals N_{ji} in contact with the individuals of the neighboring subpopulation i. In practice, this is similar to separating the commuting time scale from the other time scales in the problem (disease dynamics, traveling dynamics, etc.). While the approximation holds exactly only in the limit $\tau \to \infty$, it is good

enough as long as τ^{-1} is much smaller than the typical transition rates of the disease dynamics. In the case of ILIs, the typical time scale separation between τ and the compartments transition rates is close to one order of magnitude or even larger. The Eq.s [4.17] can be then generalized in the time scale separation regime to all compartments $[m]$ obtaining the general expression

$$X_{jj}^{[m]} = \frac{X_j^{[m]}}{(1+\sigma_j/\tau)} \text{ and } X_{ji}^{[m]} = \frac{X_j^{[m]}}{(1+\sigma_j/\tau)}\sigma_{ji}/\tau \ , \tag{4.17}$$

where $\sigma_j = \sum_i \sigma_{ji}$ denotes the total commuting rate of j. For all the other compartments which are restricted from traveling $X_{jj}^{[m]} = X_j^{[m]}$ and $X_{ji}^{[m]} = 0$. These expressions will be used to obtain the effective force of infection taking into account the interactions generated by the commuting flows.

4.8 Effective force of infection

The force of infection λ_j that a susceptible population of a subpopulation j sees can be decomposed into two terms: λ_{jj} and λ_{ji}. The component λ_{jj} refers to the part of the force of infection whose origin is local in j, while λ_{ji} indicates the force of infection acting on susceptibles of j during their commuting travels to a neighboring subpopulation i. The effective force of infection can be estimated by summing these two terms weighted by the probabilities of finding a susceptible from j in the different locations, S_{jj}/S_j and S_{ji}/S_j, respectively. Using the time-scale separation approximation that establishes the equilibrium populations in Eq. (4.17), we can write

$$\lambda_j = \frac{\lambda_{jj}}{1+\sigma_j/\tau} + \sum_i \frac{\lambda_{ji}\sigma_{ji}/\tau}{1+\sigma_j/\tau} \quad . \tag{4.18}$$

We will focus now on the calculation of each term of the previous expression. The force of infection occurring in a subpopulation j is due to the local infectious persons staying at j or to infectious individuals from a neighboring subpopulation i visiting j and so we can write

$$\lambda_{jj} = \frac{\beta_j}{N_j^*}\left[I_{jj}^{nt} + I_{jj}^{t} + r_\beta I_{jj}^{a} + \sum_i \left(I_{ij}^{nt} + I_{ij}^{t} + r_\beta I_{ij}^{a}\right)\right] \quad , \tag{4.19}$$

where in β_j we keep explicit the dependance from the subpopulation j and in our model it accounts for seasonality effects in the infection transmission rate by introducing a function of the calendar time (if the seasonality is not considered, it is a constant), and N_j^* stands for the total effective population in the subpopulation j. By definition, $I_{jj}^{nt} = I_j^{nt}$ and $I_{ji}^{nt} = 0$ for $j \neq i$. If we use the equilibrium values of the other infectious compartments (see Eq. (4.17)) we obtain

$$\lambda_{jj} \quad = \quad \frac{\beta_j}{N_j^*}\left[I_j^{nt} + \frac{I_j^t + r_\beta I_j^a}{1+\sigma_j/\tau} + \sum_i \frac{I_i^t + r_\beta I_i^a}{1+\sigma_i/\tau}\sigma_{ij}/\tau\right] \quad . \tag{4.20}$$

The derivation of λ_{ji} follows from a similar argument yielding:

$$\lambda_{ji} = \frac{\beta_i}{N_i^*}\left[I_{ii}^{nt} + I_{ii}^{t} + r_\beta I_{ii}^{a} + \sum_{\ell\in\upsilon(i)}\left(I_{\ell i}^{nt} + I_{\ell i}^{t} + r_\beta I_{\ell i}^{a}\right)\right] \quad , \tag{4.21}$$

where $\upsilon(i)$ represents the set of neighbors of i, and therefore the terms under the sum are due to the visits of infectious individuals from the subpopulations ℓ, neighbors of i, to i. By plugging the equilibrium values of the compartment into the above expression, we obtain

$$\lambda_{ji} \quad = \quad \frac{\beta_i}{N_i^*}\left[I_i^{nt} + \frac{I_i^t + r_\beta I_i^a}{1+\sigma_i/\tau} + \sum_{\ell\in\upsilon(i)}\frac{I_\ell^t + r_\beta I_\ell^a}{1+\sigma_\ell/\tau}\sigma_{\ell i}/\tau\right] \quad . \tag{4.22}$$

Finally, in order to have an explicit form of the force of infection we need to evaluate the effective population size N_j^* in each subpopulation j, i.e., the actual number of individuals actually staying at the location j. The effective population is $N_j^* = N_{jj} + \sum_i N_{ij}$, that in the time-scale separation approximation reads

$$N_j^* = I_j^{nt} + \frac{N_j - I_j^{nt}}{1+\sigma_j/\tau} + \sum_i \frac{N_i - I_i^{nt}}{1+\sigma_i/\tau}\sigma_{ij}/\tau \quad . \tag{4.23}$$

Note that in these equations all the terms with compartments have an implicit time dependence. By inserting λ_{jj} and λ_{ji} into Eq. (4.18), it can be seen that the expression for the force of infection includes terms of zeroth, first and second order on the commuting ratios (i.e., σ_{ij}/τ). These three term types have a straightforward interpretation: The zeroth order terms represent the usual force of infection of the compartmental model with a single subpopulation. The first order terms account for the effective contribution generated by neighboring subpopulations with two different sources: a transmission might occur between either susceptible individuals of subpopulation j having contacts with infectious individuals of neighboring subpopulations i, or infectious individuals of subpopulations i visiting subpopulation j. The second order terms correspond to an effective force of infection generated by the contacts of susceptible individuals of subpopulation j meeting infectious individuals of subpopulation ℓ (neighbors of i) when both are visiting subpopulation i. This last term is very small in comparison with the zeroth and first order terms, typically around two orders of magnitude smaller, and in general can be neglected.

5

Global spread of H1N1 pandemic influenza

Contents

In this chapter we present our works on the influenza pandemic H1N1. It is worth to notice that most of these results were achieved in real time during the actual unfolding of the epidemic. In particular we submitted the first four articles listed below, well before the epidemic peak. Some results went beyond the pure academic research, trying to suggest some important insights and forecasts for the policy making. The chapter is organized as follows. We first present our work on the estimate of the reproduction number R_0 of the H1N1 epidemic based on knowledge of human mobility patterns. We use GLEaM (78; 114) to simulate the worldwide evolution of the pandemic and perform a maximum likelihood analysis of the reproduction number R_0 of the H1N1 influenza against the actual chronology of newly infected countries. Subsequently, a correlation analysis allows the selection of the most probable seasonal behavior based on the observed pattern, leading to the identification of plausible scenarios for the unfolding of the pandemic and the estimate of pandemic activity peaks in the different hemispheres. We also

study the effect of systematic therapeutic use of antiviral drugs on the epidemic timeline and give an estimation of the initial number of cases in Mexico. We then present the results on a hypothetical massive vaccination campaign as it was devised in the fall 2009. We finally assess the impact of travel related measures, demonstrating both from a computational and theoretical point of view their scarce efficacy. The last section is devoted to compare and discuss the numerical results of GLEaM with the data became available after the epidemic wave.
The results presented here are based on the following papers:

- *Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility* published in BMC medicine in September 2009,
- *Estimate of Novel Influenza A/H1N1 cases in Mexico at the early stage of the pandemic with a spatially structured epidemic model* published in PLoS current Influenza in November 2009,
- *Modeling vaccination campaigns and the fall/winter 2009 activity of the new A(H1N1) influenza in the northern hemisphere* published in Emerging Health Threats Journal in November 2009,
- *Modeling the critical care demand and antibiotics resources needed during the fall 2009 wave of influenza A(H1N1) pandemic* published in PLoS current Influenza in December 2009,
- *Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic* published in PLoS One in January 2011,
- *GLEaM, a Global Stochastic Simulation Model of Influenza Epidemic: Its Application to the 2009 A/H1N1pdm* in preparation.

5.1 Background

Beginning April, 2009, the world experienced its latest global pandemic outbreak, originated in Mexico. It spread quickly to many countries in months and on June 11, 2009, the World Health Organization officially raised the phase of pandemic alert to level 6. As of July 19, 2009, 137,232 cases of the new H1N1 influenza strain had been officially confirmed in 142 different countries, and during the summer of 2009, the pandemic unfolding in the Southern hemisphere was under

scrutiny to gain insights about the succesive winter wave in the North. A major challenge was given by the need to estimate the virus transmission potential and to assess its dependence on seasonality aspects in order to use numerical models capable to forecast the spatio-temporal pattern of the pandemic.

Estimating the transmission potential of a newly emerging virus is crucial when planning for adequate public health interventions to mitigate its spread and impact, and to forecast the expected epidemic scenarios through sophisticate computational approaches (44; 111; 115; 116). With the 2009 outbreak of the new influenza A(H1N1) strain having reached pandemic proportions, the investigation of the influenza situation worldwide could be considered as the key for the understanding of the transmissibility observed in different regions and to the characterization of possible seasonal behavior. During the early phase of an outbreak, this task is hampered by inaccuracies and incompleteness of available information. A different detection rate might occur even within the same country. During the initial stage of the outbreak the enhanced setup of monitoring systems led to more accurate notifications, while later on it relaxed as reporting requirements changed (117). Reporting is also constrained by the difficulties in confirming large numbers of cases through specific tests and serological analysis. The cocirculation of multiple strains, the presence of asymptomatic cases that go undetected, the impossibility to monitor mild cases that do not seek health care and the possible delays in diagnosis and reporting, all worsen the situation. Early modeling approaches and statistical analysis show that the number of confirmed cases by the Mexican authorities during the early phase was underestimated by a factor ranging from one order of magnitude (118) to almost three (119). The Centers for Disease Control (CDC) in the US estimate a 5% to 10% case detection, similar to other countries facing large outbreaks, with expected heterogeneities due to different surveillance systems.

By contrast, the effort put in place by the World Health Organization (WHO) and health protection agencies worldwide provided an unprecedented amount of data and, at last, the possibility of following in real time the pandemic chronology on the global scale. In particular, the border controls and the enhanced surveillance aimed at detecting the first cases reaching uninfected countries appeared to provide more reliable and timely information with respect to the raw count of cases at the local level. Moreover, data on international passenger flows from Mexico was found to display a strong correlation with confirmed H1N1 importations from Mexico (120).

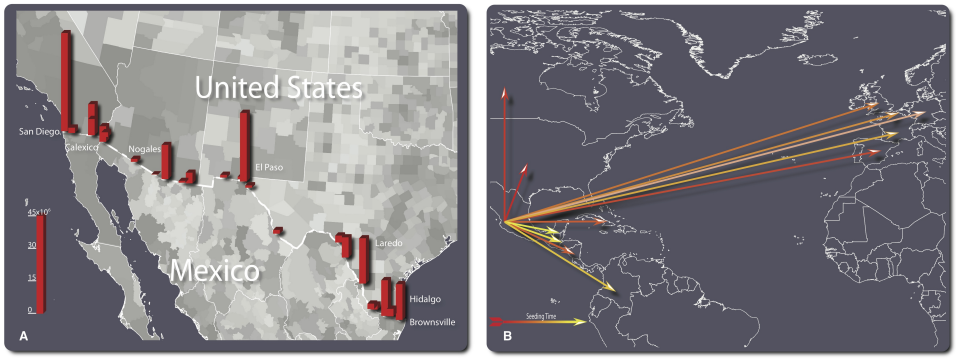

Figure 5.1: Illustration of the model's initialization. (A) Intensity of the commuting between US and Mexico at the border of the two countries. (B) The 12 countries infected by international travelers coming from Mexico used in the Monte Carlo likelihood analysis. The color scale of the arrows from red to yellow indicates the time ordering of the epidemic invasion.

5.2 Disease parameters estimation

We used the classic influenza-like illness compartmentalization in which each individual is classified by a discrete state such as susceptible, latent, infectious symptomatic, infectious asymptomatic or permanently recovered/removed (71; 121). The model therefore assumes that the latent period is equivalent to the incubation period and that no secondary transmissions occur during the incubation period (see Figure 4.2) for a detailed description of the compartmentalization). As explained in the previous chapter all transitions are modeled through binomial and multinomial processes to preserve the discrete and stochastic nature of the processes. Asymptomatic individuals are considered as a fraction $p_a = 33\%$ of the infectious individuals (101) generated in the model and assumed to infect other individuals with a relative infectiousness of $r_\beta = 50\%$ (118; 121; 122). Change in traveling behavior after the onset of symptoms is modeled with the probability $1 - p_t$, set to 50%, that individuals stop traveling when ill (121). As discussed in chapter 3, the spreading rate of the disease is ultimately governed by the basic reproduction number R_0. By computing the eigenvalues of the Jacobian at the disease-free equilibrium (76), we obtained the following expression for the basic reproductive number of the adopted compartmentalization:

$$R_0 = \frac{\beta}{\mu}(r_\beta p_a + 1 - p_a) \tag{5.1}$$

Once the disease parameters and initial conditions based on available data are defined, GLEaM allows the generation of stochastic realizations of the worldwide unfolding of the epidemic, with

mobility processes entirely based on real data. The model generates in silico epidemics for which we can gather information such as prevalence, morbidity, number of secondary cases, number of imported cases and other quantities for each subpopulation and with a time resolution of 1 day. While global models are generally used to produce scenarios in which the basic disease parameters are defined from the outset, here we use the model to provide a maximum likelihood estimate of the transmission potential by finding the set of disease parameters that best fit the data on the arrival time of cases in different countries worldwide. The projections for the winter season in the northern hemisphere were also assuming that there would not be mutation of the virus with respect to the spring/summer of 2009. Furthermore, while at the moment of our analysis the novel H1N1 influenza was accounting for 75% of the influenza cases worldwide, the model did not consider the cocirculation of different influenza strains and cannot provide information on cocirculation data.

The initial conditions of the epidemic were defined by setting the onset of the outbreak near La Gloria in Mexico on February 18 2009, as reported by official sources (123) and analogously to other works (118). We tested different localizations of the first cases in census areas close to La Gloria without observing relevant variations with respect to the observed results. We also performed sensitivity analysis on the starting date by selecting a seeding date anticipated or delayed by 1 week with respect to the date available in official reports (123). The arrival time of infected individuals in the countries seeded by international travelers departed from Mexico depends both on the number of cases present in the originating country (Mexico) and the mobility network, both within Mexico and connecting Mexico with countries abroad. For this reason we integrated into our model the data on Mexico-US border commuting (see Figure 5.1)A, which could be relevant in defining the importation of cases in the US, along with Mexican internal commuting patterns that were responsible for the diffusion of the disease from rural areas such as La Gloria to transportation hubs such as Mexico City. In addition, we used a time-dependent modification of the reproductive number in Mexico as in (119) to model the control measures implemented in the country starting April 24 and ending May 10, as those could have affected the spread to other countries. In order to ascertain the effect of seasonality on the observed pattern, we explored different seasonality schemes. The seasonality is modeled by a standard forcing that rescales the value of the basic reproductive number into a seasonally rescaled reproductive number, $R(t)$, depending on time. The seasonal rescaling is time and location dependent by means of a scaling multiplicative factor generated by a sinusoidal function with a total period of 12 months oscillating in the range α_{min} to α_{max}, with $\alpha_{max} = 1.1$ (sensitivity analysis in the

range 1.0 to 1.1) and α_{min} a free parameter to be estimated (124). The rescaling function is in counter phase in the Northern and Southern hemispheres, so that during the warmer seasons R_0 is reduced and the disease transmissibility is hampered. No rescaling is assumed in the equatorial area between the Tropics. The value of R_0 reported in the Table 5.2 and the definition of the baseline is the reference value in the Tropics. In each subpopulation the $R(t)$ relative to the corresponding geographical location and time of the year is used in the simulations.

We have defined a Monte Carlo likelihood analysis for the assessment of the seasonal transmission potential of the 2009 A(H1N1) influenza based on the analysis of the chronology of case detection in affected countries at the early stage of the epidemic. This method allows the use of data coming from the border controls and the enhanced surveillance aimed at detecting the first cases reaching uninfected countries. This data is, in principle, more reliable than the raw count of cases provided by countries during the evolution of the epidemic. The procedure provided the necessary input to the large-scale computational model for the analysis of the unfolding of the pandemic. The seasonal transmission potential of the H1N1 strain was assessed in a two-step process that first estimated the reproductive number in the Tropics region, where seasonality is assumed not to occur, by focusing on the early international importation of infective cases from Mexico, and then estimated the degree of seasonal damping factor by examining a longer time period of international spread to allow for seasonal changes. The estimation of the reproductive number was performed through a maximum likelihood analysis of the model fitting the data of the early chronology of the H1N1 epidemic. As shown in equation 5.1, the basic reproductive number R_0 is a function of r_β and p_a that are fixed according to previous clinical estimates, while the values β and μ are the free variables of the fitting procedure. In particular we explore some values of the recovery rate μ and we perform a maximum likelihood estimates by systematically varying β. Given a set of values of the disease parameters, we produced $2 \cdot 10^3$ stochastic realizations of the pandemic evolution worldwide for each R_0 value. Our model explicitly takes into account the classes of symptomatic and asymptomatic individuals and allows the tracking of the importation of each symptomatic individual and of the onset of symptoms of exposed individuals transitioning to the symptomatic class, as observables of the simulations. This allows us to obtain numerically with a Monte Carlo procedure the probability distribution $P_i(t_i)$ of the importation of the first infected individual or the first occurrence of the onset of symptoms for an individual in each country i at time t_i. Asymptomatic individuals do not contribute to the definition of t_i. In Figure 5.2 a schematic representation of the Monte Carlo likelihood procedure is shown. With the aim of working with conditional independent variables we restricted the likelihood analysis to

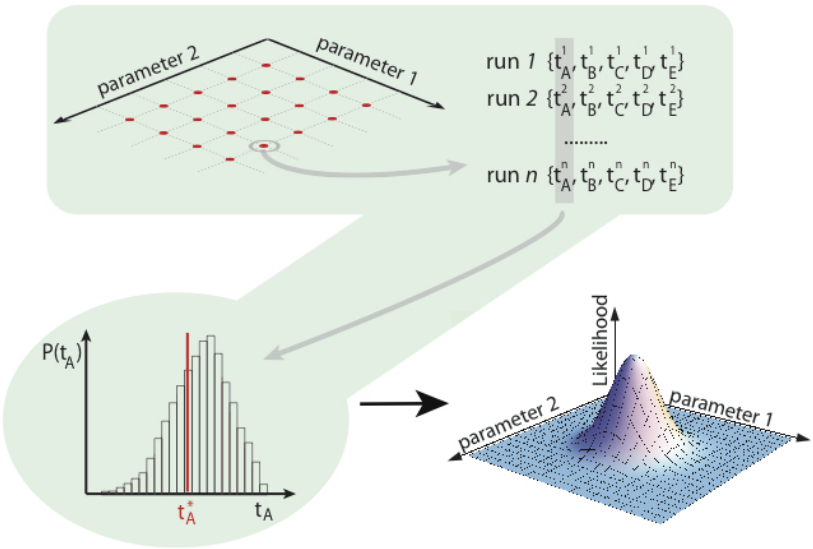

Figure 5.2: Flow chart representing the steps that compose the Monte Carlo maximum likelihood method. First, for each point in the parameter space we run $n = 2,000$ stochastic realizations, all with the same initial conditions. Second, for each run we record the arrival times of the countries under study. Third, we compare the probability distribution built on the simulated arrival times with the empirically observed arrival time for each country. Finally, we evaluate the likelihood function in order to find its maximum value, corresponding to the set of parameters that best fit the data.

12 countries seeded by infectious travelers coming from Mexico (see Figure 5.1B) and for which it was possible to know with good confidence the onset of symptoms and/or the arrival date of the first detected case (see Table 5.1). This allowed us to define a likelihood function:

$$\mathcal{L} = \prod_i P_i(t_i^*), \tag{5.2}$$

where t_i^* is the empirical arrival time from the H1N1 chronological history in each of the selected countries. Maximizing this function, after fixing the values of the epidemiological and seasonality parameters (ϵ, μ, α_{min}), we obtained an estimation of the basic reproductive number. This methodology assumes the prompt detection of symptomatic cases at the very beginning of the outbreak in a given country, and for this reason we have also provided a sensitivity analysis accounting for a late/missed detection of symptomatic individuals as reported in the next section. The transmission potential was estimated as the value of R_0 that maximizes the likelihood function $\mathcal{L}$, for a given set of values of the disease parameters. In Table 5.2 we report the reference values assumed for some of the model parameters and the range explored with the sensitivity analysis. At the time of our study there were no precise clinical estimates of the basic model parameters ϵ and μ defining the inverse average exposed and infectious time durations (125; 126; 127). The generation interval $G_t = \epsilon^{-1} + \mu^{-1}$ (128; 129), which is the mean time interval between the infection of one person and the infection of the people that this individual infects, is based on the early estimate of (118) and values obtained for previous pandemic and seasonal influenza (101; 111; 121; 122; 130; 131), with most studies focusing on values ranging from 2 to 4 days (118; 132; 133; 134). We have therefore assumed a short exposed period value $\epsilon^{-1} = 1.1$ as indicated by early estimates (118) and compatible with recent studies on seasonal influenza (101; 135) and performed a sensitivity analysis for values as large as $\epsilon^{-1} = 2.5$ days and we systematically explore the average infectious period μ^{-1} in the range $1.1 - 4.0$. The major problem in the case of projections on an extended time horizon is the seasonality effect that in the long run is crucial in determining the peak of the epidemic. In order to quantify the degree of seasonality observed in the 2009 epidemic, we estimated the minimum seasonality scaling factor α_{min} of the sinusoidal forcing by extending the chronology under study and analyzing the whole data set composed of the arrival dates of the first infected case in the 93 countries affected by the outbreak as of June 18. The full exploration of the phase space of epidemic parameters and seasonality scenarios required data from 10^6 simulations; the equivalent of 2 million minutes of PowerPC 970 2.5 GHz CPU time.

Country	Onset of symptoms	Flight arrival	Confirmed on
United States	March 28 (136)	–	April 21 (136)
Canada	April 11 (137)	April 8 (138)	April 23 (139)
El Salvador	–	April 19 (140)	May 3 (141)
United Kingdom	April 24 (142)	April 21 (143)	April 27 (139)
Spain	April 25 (144)	April 22 (145)	April 27 (139)
Cuba	–	April 25 (146)	May 13 (139)
Costa Rica	April 25 (147)	April 25 (147)	May 2 (139)
Netherlands	–	April 27 (148)	April 30 (148)
Germany	April 28 (149)	–	April 29 (139)
France	–	–	May 1 (150)
Guatemala	May 1 (151)	–	May 5 (152)
Colombia	–	–	May 3 (153)

Table 5.1: The day of onset of symptoms, flight arrival and day of official confirmation of the first confirmed case in 12 countries seeded by infectious individuals traveling from Mexico are reported.

Parameter	Best Estimate	95% CI	Description
R_0	1.75	1.64 to 1.88	Basic reproduction number
α_{min}	0.65	0.6 to 0.7	winter minimal seasonality rescaling

	Assumed values	Sensitivity analysis range	Description
μ^{-1}	2.5	1.1 to 4.0	Mean infectious period (days)
ϵ^{-1}	1.1	1.1 to 2.5	Mean exposed period (days)
α_{max}	1.1	1.0 to 1.1	Summer maximum seasonality rescaling

Table 5.2: Best Estimates of the epidemiological parameters. Estimates from the Monte Carlo likelihood analyses for various values of the parameter space explored. The confidence interval is determined by the likelihood procedure.

Table (5.2) reports the results of the maximum likelihood procedure and of the correlation analysis on the arrival times for the estimation of α_{min}. In the following we consider as the baseline case the set of parameters $G_t = 3.6$ days, $\mu^{-1} = 2.5$ days, and the best estimates $R_0 = 1.75$. The best estimate for R_0 was higher than the one obtained in early findings but close to subsequent analysis on local outbreaks (132; 133; 134). The R_0 we report is the reference value for Mexico and the tropical region, whereas in each country we have to consider the $R(t)$ due to the seasonality rescaling depending on the time of the year, as shown in Figure 5.4. This might explain the lower values found in some early analysis in the US. The transmission potential emerging from our analysis was close to estimates for previous pandemics (91; 154). We performed a full sensitivity analysis concerning the assumptions used in the model (96). Results show that larger values of the generation interval provide increasing estimates for R_0. Fixing the latency period to $\epsilon^{-1} = 1.1$ days and varying the mean infectious period in the plausible range 1.1 to 4.0 days yields corresponding maximum likelihood estimates for R_0 in the range 1.4 to 2.1. Variations in the latency period from $\epsilon^{-1} = 1.1$ to $\epsilon^{-1} = 2.5$ days provide corresponding best estimates for R_0 in the range 1.9 to 2.3, if we assume an infectious period of 3 days. We tested variations of the compartmental model parameters p_a, and p_t up to 20% and explored the range $r_\beta = 20\%$ to 80%, and sensitivity on the value of the maximum seasonality scaling factor α_{max} in the range 1.0 to 1.1. The obtained estimates lie within the confidence intervals of the best estimate values.

The empirical arrival time data used for the likelihood analysis are necessarily an overestimation of the actual date of the importation of cases as cases could go undetected. By considering earlier arrival times assuming a shift of 7 days from official reports, the resulting maximum likelihood yields an increase to the best estimate for R_0 to 1.87 (95% CI 1.73 to 2.01), as expected since earlier case importation necessitates a larger growth rate of the epidemic. The official timeline used here therefore provided, all other parameters being equal, a lower estimate of the transmission potential. We have also explored the use of a subset of the 12 countries, always generating results within the confidence interval of the best estimate.

The best estimates reported in Table 5.2 do not show any observable dependence on the assumption about the seasonality scenario. The analysis is restricted to the first countries seeded by infected individuals departing from Mexico to preserve the conditional independence of the variables and it is natural to see the lack of any seasonal signature since these countries receive the disease from a single country, mostly found in the tropical region where no seasonal effects are expected.

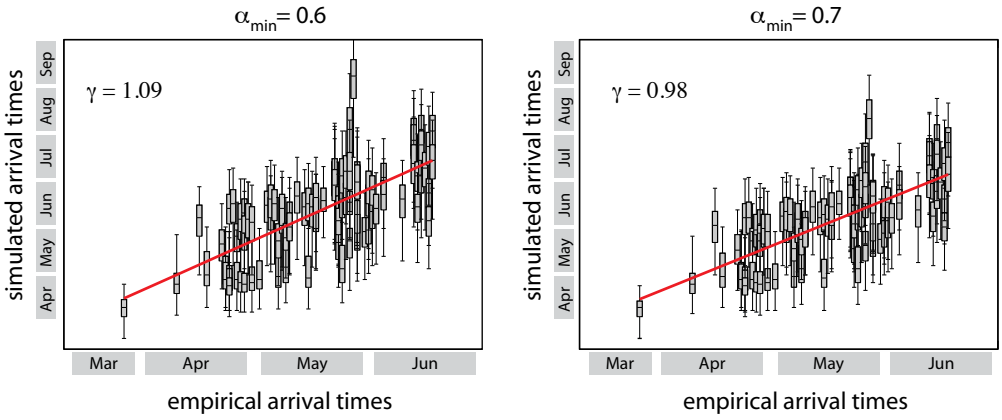

Figure 5.3: Simulated arrival times median and 95% CI versus the empirical ones for the two seasonality scaling factor 0.6 and 0.7.

In order to find the minimum seasonality scaling factor α_{min} that best fits the empirical data, we performed a statistical correlation analysis of the arrival time of the infection in the 93 countries infected as of June 18. Given the extended time frame under observation, the arrival times considered in this case are expected to provide a signature of the presence of seasonality. They included the seeding events of new countries from outbreaks taking place in regions where seasonal effects might occur, such as for example in the US or in the UK. For the simulated arrival times we have considered the median and 95% confidence interval (CI) emerging from the $2 \cdot 10^3$ stochastic runs. By considering a larger number of countries and a longer period for the unfolding of the epidemic worldwide as seasons change, the correlation analysis for the baseline scenario provides clear statistical indications for a minimum rescaling factor in the interval $0.6 < \alpha_{min} < 0.7$ as shown in Figure 5.3. We analyzed the correlation between the simulated arrival time and its corresponding empirical value, by measuring the regression coefficient (slope γ) between the two datasets. In the full range of epidemic parameters explored, the correlation analysis yields values for α_{min} in the range 0.4 to 0.9. This evidence for a mild seasonality rescaling is consistent with the activity observed in the months of June and July in Europe and the US where the epidemic progression has not stopped and the number of cases kept increasing considerably (see also Figure 5.4) for the corresponding values of $R(t)$ in those regions during summer months).

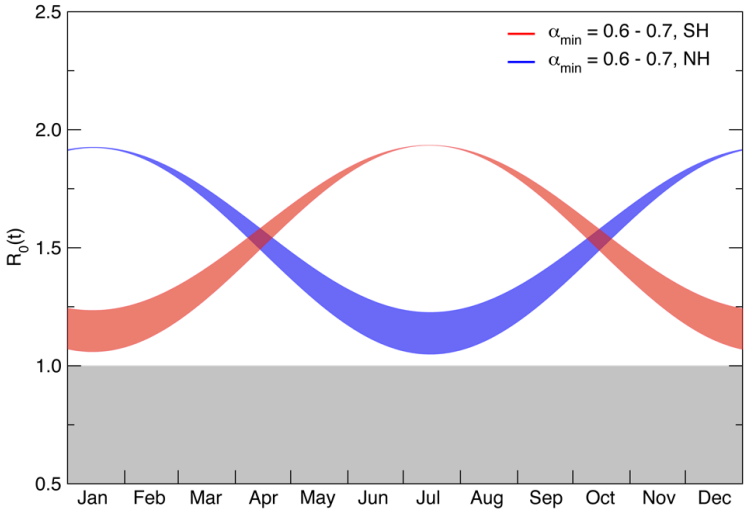

Figure 5.4: Seasonality time-dependent reproduction number in the Southern hemisphere (red shaded area) and in the Northern hemisphere (blue shaded area). The values of $R(t)$ for the Northern hemisphere correspond to the rescaling of the maximum likelihood value of R_0 in Mexico and in the Tropical regions ($R_0 = 1.75$) and the best values for the seasonality rescaling factor, $0.6 < \alpha_{min} < 0.7$. The parameter α_{min} indicates the minimum value of the seasonal rescaling of R_0 induced by the sinusoidal forcing in the Northern hemisphere (124).

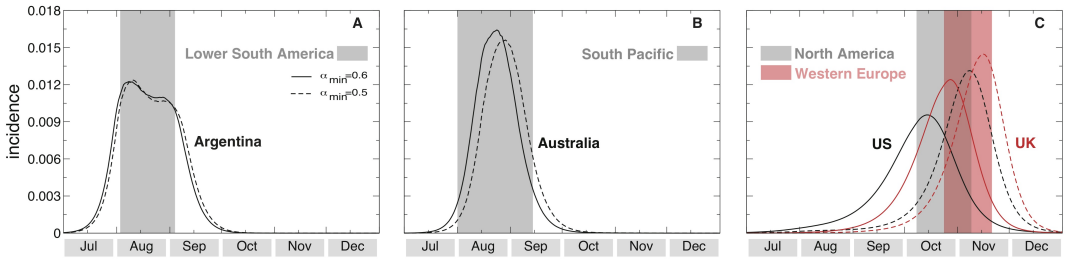

Figure 5.5: Results for the activity peaks in three geographical areas. Panels (A), (B) and (C) show the daily incidence in Lower South America, South Pacific and North America/Western Europe, respectively. The shaded area indicates the 95% confidence interval (CI) of the peak time in the corresponding geographical region. The median incidence profiles of selected countries are shown for the two values defining the best-fit seasonality scaling factor interval.

5.3 Real time predictions

After the epidemic parameters estimation we were able to provide a comparison with the epidemic activity observed and an early assessment of the subsequent unfolding of the epidemics. For each set of parameters the model generates quantities of interest such as the profile of the epidemic behavior in each subpopulation or the number of imported cases. Each simulation generates a stochastic realization of the process and the curves are the statistical aggregate of at least $2 \cdot 10^3$ realizations. In the following we report the median profiles and where indicated the 95% CI. Results are in good agreement with the reported temporal evolution of the epidemic and highlight a progressive decrease of the monitoring activity caused by the increasing number of cases, as expected (117).

In Figure 5.5A-B we report the real time predictions of the baseline case for countries in the Southern hemisphere. It is possible to observe in the figure that in this case, the effect of seasonality is not discriminating between different waves, as the short time interval from the start of the outbreak to the winter season in the Southern hemisphere does not allow a large variation in the rescaling of the transmissibility during these months. Therefore we predicted a first wave occurring between August and September in phase with the seasonal influenza pattern, and independently of the seasonality parameter α_{min}. The situation was expected to be different in the Northern hemisphere where different seasonality parameters would have progressively shifted the peak of the epidemic activity in the winter months. Figure 5.5C reports the predicted daily incidence profiles for the Northern hemisphere and the 95% CI for the activity peaks of the pandemic

with the best-fit seasonality scenario (that is, the range $0.6 < \alpha_{min} < 0.7$). The general evidence clearly points to the occurrence of an autumn/winter wave in the Northern hemisphere strikingly earlier than usual, with peak times ranging from early October to the middle of November while the seasonal influenza usually peaks in January-February. The peak estimates for each geographical area is obtained from the epidemic profile summing up all subpopulations belonging to the region. The activity peak estimate for each single country can be noticeably different from the overall estimate of the corresponding geographical region as more populated areas may dominate the estimate for a given area. For instance Chile had a pandemic activity peak in the interval July 1 - August 6, one month earlier than the average peak estimate for the Lower South America geographical area it belongs to. It is extremely important to remark that in the whole phase space of parameters explored the peak time for the epidemic activity in the Northern hemisphere lay in the range from late September to late November, thus suggesting that the early seasonal peak was a genuine feature induced by the epidemic data available at the moment of the analysis.

In order to assess the amount of pressure on the healthcare infrastructure, we provided the expected number of hospitalizations at the epidemic peak according to different hospitalization rate estimates (see Table 5.3) . The assessment of the hospitalization rate was very difficult as it depends on the ratio between the number of hospitalizations and the actual number of infected people. As discussed previously, the number of confirmed cases released by official agencies was always a crude underestimate of the actual number of infected people. We considered three different methods along the lines of those developed for the analysis of fatalities due to the new virus (155). The first assumes the average value of hospitalization observed during the regular seasonal influenza season. The second is a multiplier method in which the hospitalization rate was obtained as the ratio between the WHO number of confirmed hospitalizations and the cases confirmed by the WHO multiplied by a factor 10 to 30 to account for underreporting. The third method is given by the ratio of the total number of confirmed hospitalizations and the total number of confirmed cases. This number was surely a gross overestimation of the hospitalization rate (155; 156). It has to be noted that hospitalizations were often related to existing health conditions, age and other risk factors. This implies that hospitalizations would likely have not affected the population homogenously, a factor that we cannot consider in our model.
The number of hospitalized at peak times in the selected countries range between 2 and 40 per 100.000 persons, for a hospitalization rate typical of seasonal influenza and for an assumed 1% rate, respectively, yielding a quantitative indication of the potential burden that the health

	Seasonal influenza	Multiplier method		WHO confirmed cases
	HR:0.08%	HR:0.3%	HR:1%	HR:10%
USA	2.21	8.28	27.58	275.84
Canada	2.18	8.17	27.22	272.23
UK	2.52	9.45	31.52	315.15
France	2.61	9.79	32.64	326.40
Germany	2.98	11.17	37.22	372.18
Italy	72.87	10.76	35.87	358.67
Spain	2.54	9.54	31.81	318.12
China	2.48	9.32	31.05	310.50
Japan	2.59	9.70	32.32	323.19

Table 5.3: Number of hospitalizations per 100,000 persons at the activity peak in several countries. The estimates are obtained by considering three methods. The first assumes the average hospitalization rate (HR) observed during the seasonal influenza season. The second is a simple multiplier method in which the HR is obtained as the ratio between the World Health organization (WHO) number of confirmed hospitalizations and the cases confirmed by the WHO multiplied by a factor 10 to 30 to account for underreporting. The third method is simply the ratio of the total number of confirmed hospitalizations and the total number of confirmed cases.

care systems would have faced at the peak of the epidemic activity in the fall of 2009. It is worth noting that the present analysis considered a worst-case scenario in which no effective containment measures were introduced. This was surely not the case in that pandemic plans and mitigation strategies were considered at the national and international level. Guidelines aimed at increasing social distancing and the isolation of cases were crucial in trying to mitigate and delay the spread in the community, thus reducing the overwhelming requests on the hospital systems.

5.4 Estimating the early number of cases in Mexico

By using GLEaM it was possible to provide a model estimate of the number of imported cases arriving from Mexico to a set of selected countries. The estimated 99% reference range is shown in Table 5.4. The dates and target countries were chosen to facilitate the comparison with the numbers found in the literature (157; 158; 159; 160). The numbers shown in the Table refer to the importation of infected/exposed individual traveling from Mexico in one of the listed countries as of the date of May 8^{th}. Only 2/3 of the exposed travelers were then considered in the cumulative number of cases as only this fraction will eventually develop symptoms, according to the model

assumptions. The numbers of imported cases to each country were typically small, and as such prone to large stochastic fluctuations. However the surveillance values were all within the 99% reference ranges of the $2 \cdot 10^3$ realizations of our model. We provided in Ref. (96) a full sensitivity analysis of the results but, overall, there are very small variations with respect to the presented results in the range of parameters explored. This is because any Maximum Likelihood Estimate (MLE) for R_0 and generation interval tends to optimize the growth rate with respect to the epidemic timeline thus producing very similar results in the early spreading of the epidemic. We have also considered that in the US the travel history was known only for 50% of the confirmed cases. The simple extrapolation that provides a twofold estimate of imported cases (in brackets in Table 5.4) was however still compatible with the reference range of our stochastic simulations. Table (5.5) shows GLEaM predictions for the size of the epidemic in Mexico on April 30 and compares the results with the estimations of Refs. (118) and (157). We provide the 95% reference range over $2 \cdot 10^3$ realizations. The obtained range included the lower bound estimate of Ref. (157). Our median value for the number of asymptomatic cases is 734.000 that is again compatible with the range of values reported in Ref.(157). While the estimates presented in Refs. (118) and (157) are based on a homogeneous mixing approach within the entire country of Mexico, the approach used here is a spatially structured model that just in Mexico counts 65 different census areas. These census areas are not equally connected internationally and between them. The number of cases relevant for the international spread of infected individuals are mostly in census areas close to international transportation hubs. Poorly connected regions of Mexico on the other hand, while experiencing a considerable number of cases, would contribute only marginally to the International spread of cases. This observation readily explains why single population calculations that matched the detection of imported cases with the local prevalence were necessarily underestimating the latter quantity.

While GLEaM takes into account a higher level of geographical organization than previous approaches, its estimates still contain a number of assumptions and approximations. The contagion within each census area is approximated by means of a homogeneous mixing process. Once a person arrives at a census area by plane, he/she becomes integrated into the local population. This implies that, as in (157), the travelers and the local population are equally exposed to the disease. Finally, the model considers each individual as independent and the possibility of cluster cases, i.e. a group of people that are in close contact and can therefore easily transmit the disease to each other, is not considered. Despite these shortcomings and other necessary uncertainties,

Number imported cases (May 8)	USA	UK	France	Germany	Brazil
Simulation Results	0 - 534	0 - 44	0 - 62	0 - 55	0 - 45
Surveillance data	85 (170)	17	11	9	3

Table 5.4: Cumulative number of imported cases from Mexico shown as the 99% reference range over $2 \cdot 10^3$ realizations on May 8 for a few countries. The simulations are obtained with the best estimate parameters of the baseline case of Ref. (96) and R_0=1.75 [95%CI 1.64 to 1.88]. The number of imported infected individuals and of independent clusters correspond to the data given in Ref. (157) for US, and UK and the values in (159) for France, in (158) for Germany and in (160) for Brazil. No data was available to assess the possible presence of clusters in Germany and France. In the USA we report in parentheses the revised number considering the rate of unknown travel history in confirmed cases.

	Number of symptomatic cases in Mexico (April 30)
Simulation Results	[121,000 - 1,394,000]
Lower bound range of Ref. (157)	113,000-375,000
Estimate of Ref. (118)	2,000 - 280,000
Mexican official report (161) (confirmed cases)	3,350

Table 5.5: Predictions of GLEaM for the size of the epidemic in Mexico on April 30 in thousands of cases and comparison with other approaches and with empirical data. The simulations are obtained with the best estimate parameters of the baseline case of Ref. (96) and show the 95% reference range over $2 \cdot 10^3$ stochastic realizations. The results are compared with the lower bound estimate range in (157), the estimate provided in Ref. (118) and the number of confirmed cases given by official reports. The interval provided for Ref.(118) is obtained by merging the results reported in the paper under different assumptions and including the 95% CI.

GLEaM predictions could provide additional information for a better understanding of the early evolution of the past pandemic. Despite the different approximations used here and in Ref.(157), both approaches were providing support to the possibility of a reporting ratio of infected cases in Mexico as low as 1 in 100, in agreement with prior estimates (119). This finding was important when evaluating the massive amount of data which were collected in a large number of countries around the world. We can easily imagine that the reporting rate as well as any estimate of the cumulative attack rate in most of the countries could be easily underestimated by orders of magnitude.

5.5 Intervention strategies

5.5.1 Vaccination campaign

As an effective line of defense against influenza epidemics most of the countries planned the vaccination of a large fraction of the population (162). The vaccine development and production started after the virus identification at the end of April 2009, and received the approval by the US Food and Drugs Administration in mid September 2009 (163). Vaccine delivery was scheduled to start in early or mid-October (163) in several countries, but the expected timing of the pandemic influenza activity predicted to peak in October/November put at risk the effectiveness of mass vaccination as a control strategy. Using GLEaM (96; 111) it was possible to assess in advance the effect of mass vaccination on the predicted pandemic evolution, given the expected vaccine availability and timing of distribution. We used the model and predicted patterns of global spread obtained in Ref.(96) to quantify the mitigation effect of mass vaccination campaigns and combined strategies under different scenarios. The baseline (no intervention) scenario was compared along with mitigation strategies based on the use of antiviral drugs and the use of vaccines (78; 115; 116; 121; 122; 130; 164; 165; 166; 167; 168). Intervention involving vaccination was constrained on the availability and distribution of vaccine doses matching the novel H1N1 influenza virus. Since this analysis has been performed in the late September-October 2009, information on the time and amount of delivery of the first doses of vaccine was available for certain countries only and underwent continuous updates. Significant availability of H1N1 vaccine was expected to begin only in mid-October or later. The United States projected to have 45M doses by October 15, with additional 15M doses shipped every week after that date, reaching the delivery of the full amount of 195M doses by the end of December (169; 170; 171). The United Kingdom planned to have the first amount of 100,000 doses by mid-October, with subsequent

distribution of additional doses till full coverage of the population (172). Little was known about vaccine production rates and delivery for several other countries. Here we assumed that all countries having stockpiled on antivirals (173) would have placed orders to have vaccines available to administer to their populations. Based on the available data on vaccination programs, we explored scenarios where the campaign started on the same date for all countries with vaccines, where the date was set to October 15 or November 15. Additional dates were also studied in the sensitivity analysis. Following previous studies on vaccination during the course of a pandemic (115; 116; 174), we assumed a dynamic mass vaccination of 1% of the population uniformly in countries where doses were available, till their exhaustion. We assume the administration of a single dose of vaccine (111; 175; 176), providing protection with a delay of 2 weeks (177). The 2 weeks time to produce the immune response was chosen according to the preliminary data in adult clinical studies for H1N1 influenza vaccine (163; 177), and a sensitivity analysis reducing it to 1 week was performed. Recommendations foresaw the use of vaccines first in the groups of population who were at elevated risk of severe outcomes or who were likely to come in contact with the novel H1N1 virus (178). The model did not consider social structure in the subpopulations, therefore the effect of prioritized distribution of vaccines to health care workers, risk groups, and others, in reducing the number of hospitalizations and deaths (178; 179; 180; 181) was out of the scope of our study. Mass vaccination aims to (i) reduce susceptibility to infection; (ii) reduce infectiousness if infection occurs; (iii) reduce the probability of developing clinical symptoms (115). The efficacy of the vaccine with respect of these three effects is quantified by the parameter VE_S, VE_I, VE_D, respectively. The efficacy of the vaccine was still under study, therefore we referred to previous estimates and performed a sensitivity analysis to explore higher and lower efficacy levels. We considered a vaccine efficacy for susceptibility $VE_S = 70\%$, a vaccine efficacy for infectiousness $VE_I = 30\%$, and a vaccine efficacy for symptomatic disease given infection $VE_D = 50\%$ (115; 179; 182). Based on the partial information on total production amounts per country, ranging from approximately 1/3 of the population (183; 184; 185) to 2/3 (169), up to full coverage (172; 186; 187), we explored two different mass vaccination scenarios in which we assumed a 30% and a 60% coverage of the population.

We also considered combined strategies including the systematic treatment of clinical cases with antiviral drugs aimed at reducing the severity of the disease and the transmissibility while infectious (121; 122; 130). The data on antiviral stockpiles in the world were collected from Ref. (173) and from national agencies to model the current availability of the drugs by country. We assumed the treatment with antivirals of 5% and 10% of clinical cases within the first day from

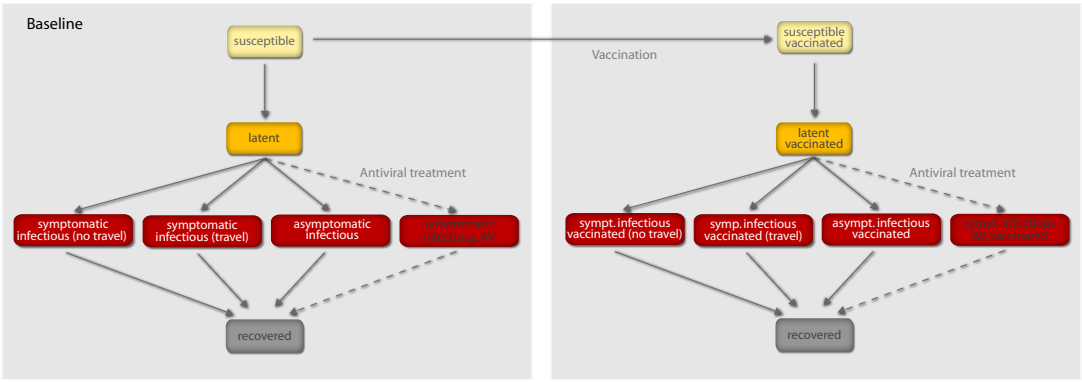

Figure 5.6: Compartmental structure in each subpopulation. A susceptible individual interacting with an infectious person may contract the illness and enter the latent compartment where s/he is infected but not yet infectious. At the end of the latency period, each latent individual becomes infectious entering the symptomatic compartment with probability $(1 - p_a)$ or becoming asymptomatic with probability p_a. Asymptomatic individuals infect with a transmission rate reduced of $r\beta$. A fraction $(1-p_t)$ of the symptomatic individuals stop traveling when ill. Infectious individuals recover permanently with rate μ. Antiviral treatment is assumed to be administered to a fraction $p_A V$ of the symptomatic infectious individuals within one day from the onset of symptoms, according to the drugs availability in the country. It reduces the infectiousness by the antiviral efficacy AVE_I and shortens the infectious period of 1 day. If vaccines are available, a fraction equal to 1% of the susceptible population enters the susceptible vaccinated compartment each day. A similar progression to the baseline compartmentalization is considered if infection occurs. However, the vaccine reduces the susceptibility of the vaccinated susceptible with an efficacy VE_S, the probability of developing symptoms if infection occurs with an efficacy VE_D, and their transmission rate while infectious with an efficacy VE_I. All transition process are modeled through multinomial processes.

the onset of symptoms, along with a hypothetical conservative intervention with the treatment of 30% of clinical cases. This parameter considered the prompt detection of symptomatic cases and the rapid administration of the drug (78; 96). The treatment was considered to last until resources were available. We assumed a drug efficacy in reducing transmission equal to 62%, and a reduction of 1 day of the total infectious period (121; 122). A schematic illustration of the compartmental diagram including the combination of intervention strategies is reported in Figure 5.6.

By using the best estimates of the model parameters as in the previous section, it was possible to calculate the 95% reference range for the activity peak in each country. The benchmark to evaluate the effect of mass vaccination campaigns was the no intervention scenario that was predicted to reach the activity peak e.g. in the United States between the beginning of October and the beginning of November. In the following we will refer to the early and late peak cases as the earliest and latest date, respectively, of the reference range for the activity peak time (96). This allows us the consideration of the whole range of peak times to explore the impact of mass vaccination campaigns also in extreme situations such as a very early activity peak in October. It is important to stress that even the "late" peak case corresponds to an activity peak occurring much earlier than the usual timing of seasonal influenza. It is also worth remarking that the predictions for the activity peak reference range obtained in the model in the Northern Hemisphere differ from country to country (96). In the case of an activity peak at the beginning of the reference range provided by the model (early October for the US and many European countries), our study predicted that the mass vaccination program starting on October 15 with 30% coverage would have almost no effect on the epidemic profile, as the effective immunization would start long after the epidemic peak. In the case of a late peak corresponding to the other extreme of the reference range (from early to late November depending on the country), the peak attack rate would be reduced by a factor of about 28% averaged across countries, ranging from 15% to 38% depending on the specific pandemic unfolding in each country, with a lower reduction obtained in those countries where the epidemic would have arrived earlier (e.g. US vs. Europe, according to our predictions). Figures 5.7 and 5.8 show the incidence curves for a set of countries in the early and late peak cases, respectively. In the US for example, the effect of mass vaccination (with no additional intervention strategy) would correspond to a 15% reduction of the peak incidence in the most favorable situation of a late peak and early vaccination campaign. If the availability of the first vaccine batches was delayed of 1 month, the mass vaccination program would have almost no mitigation effect (less than 2%) for all countries under study in

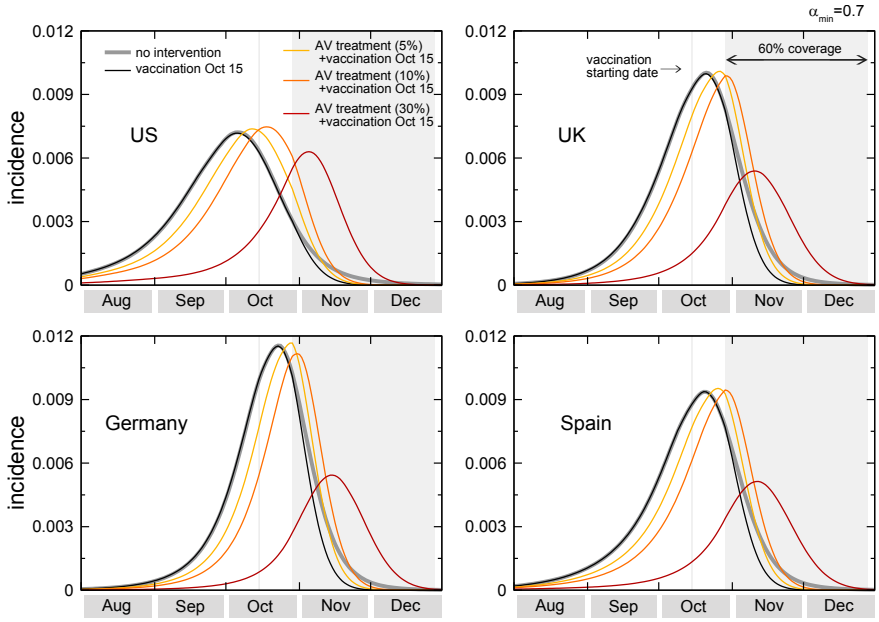

Figure 5.7: Effect of vaccination and of combined strategies for the early peak case. The incidence curves show the impact of an incremental vaccination with 1% daily distribution policy starting on October 15 for the early peak case. The baseline case was compared to the cases in which intervention strategies were considered - vaccination only, and combination of vaccination with antiviral treatment of 5%, 10%, and 30% of clinical cases. Efficacies of antiviral treatment and vaccination assumed the values reported in the text. Median profiles obtained from 2,000 stochastic realizations of the model are shown. A 60% vaccine coverage is assumed, with the gray bar indicating the time period during which the immunization takes effect.

the whole range of scenarios explored. Moreover, no major differences were observed with a larger coverage, given the 1% daily distribution rate, since in both the early and late peak extreme of the activity peak reference range the assumed 30% coverage would almost always be enough for the distribution during the entire epidemic activity, even assuming an early distribution starting on October 15. According to the simulated scenarios the mass vaccination would therefore do little against a pandemic expected to peak before or at the beginning of November, consistently with the simulation results on phased vaccination strategies in the United States (179).

We investigated if the introduction of combined mitigation strategies could help in pushing back the epidemic peak and make more effective the mass vaccination campaigns. Here we report simulations of scenarios in which the systematic use of antiviral drugs for treatment of cases is

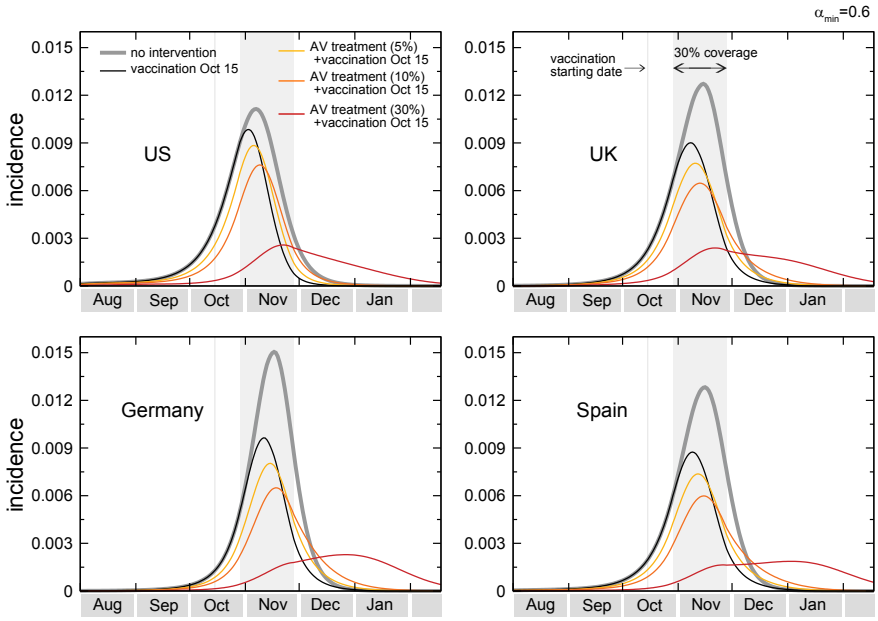

Figure 5.8: Effect of vaccination and of combined strategies for the late peak case. The incidence curves show the impact of an incremental vaccination with 1% daily distribution policy starting on October 15 for the late peak case. The baseline case was compared to the cases in which intervention strategies were considered - vaccination only, and combination of vaccination with antiviral treatment of 5%, 10%, and 30% of clinical cases. Efficacies of antiviral treatment and vaccination assume the values reported in the text. Median profiles obtained from 2,000 stochastic realizations of the model are shown. A 30% coverage is assumed, with the gray bar indicating the time period during which the immunization takes effect.

used to delay the epidemic peak, and to reduce the attack rate at peak time in combination with the vaccination campaign (78; 115; 116; 121; 122; 130; 164; 165; 166; 167; 168). If we assume a 5% to 10% detection of clinical cases and prompt administration of drugs, the pandemic peak would be delayed of approximately 1-2 weeks in the countries with available antiviral stockpiles. We also studied a possible scenario of analysis that assumed a 30% treatment, leading to approximately a full month delay of the pandemic peak (96). Though larger fraction than the implemented policy for the treatment of clinical cases in some countries, it allows the study of the effectiveness of mass vaccination campaign when a delay of one month can be achieved with a combination of intervention strategies.

Finally, it is worth noting that our model assumed a 100% susceptibility in the population, neglecting effects of prior immunity, since no clear estimates have been provided (188; 189; 190). On the other hand, the global nature of the model allows the simulation of the pandemic since its start in Mexico, taking into account the population-level immunity caused by the first peak of the spread of pandemic H1N1 in the Northern hemisphere during the spring and summer 2009. The presented results for the simulated attack rates were likely overestimating the pandemic impact because of the above assumptions. With the best estimate parameters used here, we found clinical attack rates in absence of intervention policies (i.e. baseline case) of approximately $35 - 40\%$ at the end of the epidemic. A full comparison with attack rates estimates from real data (191) was however made difficult along by the large under-ascertainment of cases, the presence of detection biases, surveillance systems with country-specific capacity and coverages, as well as monitoring requirements changing in time as the epidemic was progressing.

The interplay between the timing of the pandemic and the start of the dynamic vaccination campaign was crucial for mitigation effects. Results show that mass vaccination may have had little effect on controlling the pandemic even when administered as early as mid-October, unless additional mitigation strategies would have considered to delay the activity peak. This made also a strong case for prioritized vaccination programs focusing on high-risk groups, healthcare and social infrastructure workers. If the pandemic peak had occurred much later than anticipated from the modeling approach, in December or January, there would have been enough time to provide immunization to a larger fraction of the population given the schedule for vaccination campaign, with a larger mitigation effect than in the early pandemic wave situation.

5.5.2 Modeling the critical care demand

Following the estimates of the severity of H1N1 pandemic, we assume a complication rate of 15% of clinical cases (192), a hospitalization rate of 0.5% of clinical cases (193), and an intensive care unit (ICU) admission rate of 15% of hospitalized patients (194). We model influenza-related pneumonia as a complication associated to influenza infection, considering two main types of pneumonia: primary viral pneumonia and secondary bacterial pneumonia. While bacterial coinfection was shown to be the predominant cause of death in previous influenza pandemics (195), its presence in the severe cases analyzed since the start of the outbreak range from almost no evidence in the early reviews (196; 197; 198), to about 10% (199), 33% or larger proportions (200; 201) of the cases presenting influenza-associated complications. These fluctuations in the role of bacterial pneumonia might be due to the difficulty of testing for specific bacterial diagnosis, or to the use of antibiotics prior to routine clinical tests. Given the uncertainty on the cause of pneumonia at this stage of the epidemic evolution, we assume a proportion of bacterial pneumonia in cases showing complications in the range of $33 - 50\%$. Under pandemic conditions, it is assumed that very small differences will be implemented in the management and treatment of the patients with either types of pneumonia, as the diagnosis of influenza-associated complications will be mostly based on clinical findings and most prescribing will be empirical, based on both antibacterial therapy and antiviral medications (202). Multiple subsequent stages of pneumonia course are modeled according to a classification score which is based on clinical symptoms (203), and different progressions are assumed to take into account both viral and bacterial pneumonia. It is also worth remarking that the model does not consider social structure in the subpopulations, therefore the effect of prioritized distribution of vaccines to individuals belonging to risk groups in reducing the number of hospitalizations and deaths is not considered here.

Based on the available knowledge of complication, hospitalization and ICU rates, and the relative proportion of bacterial vs. viral pneumonia, the simulation results allow the measure of the predicted need of beds in intensive care units, and provide estimates of the corresponding courses of antibiotics needed. Figure 5.9 shows the time evolution of the predicted prevalence of ICU occupancy for a given set of countries. In the baseline case, when no intervention is implemented, the ICU prevalence peak ranges between approximately 5 and 7 ICU beds per 10^5 individuals. These values are well below the national average capacity of some countries, such as e.g. the United States with a total of about 20 ICU beds per 10^5 (204) and Germany with an average of approximately 28 ICU beds per 10^5 (205). The predicted need is slightly lowered if a

0.1% dynamic vaccination is considered, and would be reduced to values in the range of 3.6 to 4.8 ICU beds per 100,000 if we assume $r_v = 1\%$, below the national average number of ICU beds of many European countries (206). While the predicted ICU beds needs are averaged at the country level to conform with the capacity data, it is however important to note that the impact and the potential occurrence of critical situations strongly depends on the geographic distribution of the critical care resources, with areas that might have access to a larger number of intensive care units than others (see for example Ref. (207)). Moreover, a direct comparison between the simulated demand and critical care availability is made difficult by the lack of a standard definition for intensive care unit beds, and the large variations observed in both numbers of beds and volume of admission between countries in North America and Western Europe (206).
The results shown in Figure 5.9 are based on an average ICU length of staying equal to $L_{ICU} = 7$ days. Since there is a large variation in this parameter, with cohort studies showing median duration of 7 days and interquartile range up to approximately 2 weeks (201), we also explored the effect of considering longer lengths of staying, $L_{ICU} = 10$ and $L_{ICU} = 14$ days.

The longer bed occupancy would inevitably lead to an increase in the need of ICU beds at peak, in the range of approximately 9 to 12 per 100,000 persons in the case of 14 days of average ICU duration (see Table 5.6).

Along with anecdotal reports indicating ICUs overwhelmed by the sudden surge of H1N1 cases with severe complications (208), studies on the winter experience in the Southern Hemisphere during the H1N1 pandemic wave confirmed a substantial impact on ICUs, with the maximum number of ICU beds occupied by region in Australia and New Zealand ranging between 0.63 and 1.1 per 100,000 inhabitants (201). These values were smaller than the ICU demands predicted for the fall wave in the Northern Hemisphere. It is important to note, however, that the used model does not take into account the population structure (age dependent attack rates), risk groups and prior immunity thus likely overestimating the global attack rate of the pandemic. Furthermore the model did not include mitigation factors (e.g. social distancing, targeted school closures, etc.) that might have contributed to the reduction of the overall burden on the critical care facilities.

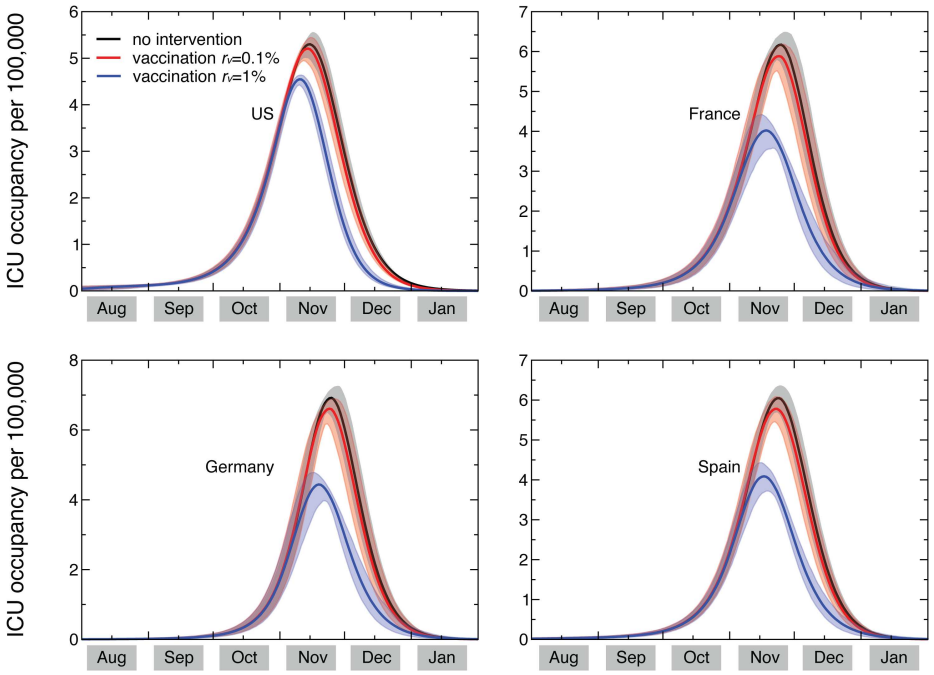

Figure 5.9: Time evolution of the ICU occupancy in a set of countries. ICU occupancy measures the predicted need of ICU beds per 100,000 persons. Results for the United States, France, Germany, and Spain are shown. The three profiles per each country refer to the predicted ICU occupancy in the baseline case when no intervention is implemented, and in case dynamic vaccination campaigns with distribution rates $r_v = 0.1\%$ and $r_v = 1\%$ are considered. Solid curves correspond to the median profiles and the shaded areas to the 95% reference range obtained from 2,000 stochastic simulations. The average ICU length of staying is assumed equal to 7 days (201).

ICU occupancy at peak (per 100,000)									
Country	**Baseline**			**Vaccination campaigns**					
				0.1%			1%		
	7 days	10 days	14 days	7 days	10 days	14 days	7 days	10 days	14 days
US	[5.0-5.6]	[6.8-7.5]	[8.7-9.7]	5.0-5.5]	[6.7-7.3]	[8.6-9.4]	[4.5-4.6]	[5.9-6.2]	[7.6-7.9]
UK	[5.7-6.5]	[7.6-8.6]	[9.9-11.0]	[5.5-6.2]	[7.4-8.2]	[9.6-10.5]	[3.9-4.6]	[5.2-6.1]	[6.7-7.7]
Canada	[5.0-5.7]	[6.7-7.6]	[8.7-9.9]	[4.8-5.5]	[6.5-7.3]	[8.5-9.5]	[3.8-4.4]	[5.1-5.8]	[6.5-7.3]
France	[5.9-6.6]	[7.9-8.7]	[10.2-11.2]	[5.7-6.2]	[7.6-8.3]	[9.8-10.6]	[3.6-4.4]	[4.9-5.9]	[6.3-7.4]
Italy	[6.5-7.1]	[8.6-9.4]	[11.0-12.0]	[6.2-6.7]	[8.2-8.9]	[10.5-11.3]	[3.6-4.5]	[4.8-5.9]	[6.1-7.4]
Spain	[5.8-6.4]	[7.8-8.6]	[10.0-11.0]	[5.6-6.1]	[7.5-8.2]	[9.6-10.5]	[3.8-4.5]	[5.1-5.9]	[6.5-7.5]
Germany	[6.6-7.3]	[8.8-9.7]	[11.2-12.2]	[6.4-7.0]	[8.5-9.2]	[10.8-11.6]	[4.0-4.8]	[5.4-6.4]	[6.8-8.0]

Table 5.6: Predicted need of ICU beds in the baseline case scenario and in the case of vaccination campaigns. The 95% reference range (RR) of the daily number of occupied ICU beds per 100,000 is reported at its peak for several countries in the Northern Hemisphere.

5.5.3 Travel restrictions

After the emergence of the H1N1 influenza in 2009, some countries responded with travel-related controls during the early stage of the outbreak in an attempt to contain or slow down its international spread. These controls along with self-imposed travel limitations contributed to a decline of about 40% in international air traffic to/from Mexico following the international alert. However, no containment was achieved by such restrictions and the virus was able to reach pandemic proportions in a short time. When gauging the value and efficacy of mobility and travel restrictions it is crucial to rely on epidemic models that integrate the wide range of features characterizing human mobility and the many options available to public health organizations for responding to a pandemic. In this section we present a comprehensive computational and theoretical study of the role of travel restrictions in halting and delaying pandemics by using a model that explicitly integrates air travel and short-range mobility data with high-resolution demographic data across the world and that is validated by the accumulation of data from the 2009 H1N1 pandemic. We explore alternative scenarios for the 2009 H1N1 pandemic by assessing the potential impact of mobility restrictions that vary with respect to their magnitude and their position in the pandemic timeline. We provide a quantitative discussion of the delay obtained by different mobility restrictions and the likelihood of containing outbreaks of infectious diseases at their source, confirming the limited value and feasibility of international travel restrictions. These results are rationalized in the theoretical framework characterizing the invasion dynamics of the epidemics at the metapopulation level.

Figure 5.10 summarizes the simulation's reproduction of the observed relative magnitude of imported cases in the local epidemics of newly-affected countries that validate the model. Panels A, B show cases in the United Kingdom and Germany, respectively, during the early phase of the outbreak when case-based surveillance was deployed in order to detect imported H1N1 cases and monitor local H1N1 transmission (209; 210). Computer simulations also allow us to explore the level of stochasticity associated with the importation of infectious individuals. We keep track for each time step of each realization of the contribution of imported cases to the total prevalence in the country defined as the ratio Q of the number of imported cases versus the total number of infectious individuals in the country. Since at the early stage of the epidemic there are usually large fluctuations in the number of imported local transmission cases, we measure the probability in time of observing a given ratio Q by averaging over 2,000 realizations of the global simulation. Panels 5.10C, 5.10D show the time behavior of the probability distribution

$P(Q)$ clearly illustrating that the importation of cases dominates the initial phase of the epidemic in each country, which is soon followed by a sustained local transmission. The contribution of imported cases is observed to reach 100% with a finite probability only during the months of April-May, after which the probability distribution progressively shrinks around small values of Q, showing how the local H1N1 transmission starts to dominate the epidemic. The agreement of the model with the actual data from the H1N1 pandemic allows us to assess the effect of the observed decline in travel flows to/from Mexico by comparing the results obtained in the reference scenario with a version of the model in which no travel reduction is considered.

Control measures based on limiting or constraining human mobility are considered in the contingency planning of several countries (211). The aim of these control measures is the decrease of travel to/from the areas affected by the epidemic outbreak and the corresponding decrease of the number of infected individuals reaching countries not yet affected by the epidemic. While the effects of slowing down the international propagation of an epidemic can be statistically evaluated based on available data and bootstrap techniques (212), the impossibility of disentangling the role played by travel from other contributing factors in the spread of an epidemic (213) has generated debates about the appropriate strategy for mobility restrictions. In this context the only way to systematically gauge uncertainty and the effectiveness of competing control strategies is through data-driven modeling efforts (111; 115; 116; 124; 214; 215). However, most previous works have focused on synthetic pandemic influenza scenarios and only a few empirical examples are available to validate models and evaluate the effectiveness of travel restrictions in general (216; 217; 218).

In the 2009 H1N1 pandemic (H1N1pdm), control measures (see Table 5.7) included travel bans to/from Mexico, the screening of travelers on entry into airports, and travel advisories against non-essential travel to Mexico (211). By considering also the spontaneous reaction of individuals to the health emergency, a reduction in the international traffic to/from Mexico of about 40% have been observed during the month of May, followed by smaller reductions in the following months, and resulting in a slow return to normality in about 3 months (229). The aggregation of data on the H1N1 pandemic therefore represents an unprecedented opportunity to calibrate and validate a modeling approach to the global spread of epidemics that integrates detailed information on human mobility and travel.

We consider as a reference scenario the one produced by the best estimates able to reproduce the initial chronology of newly infected countries (i.e. the baseline scenario), where in addition

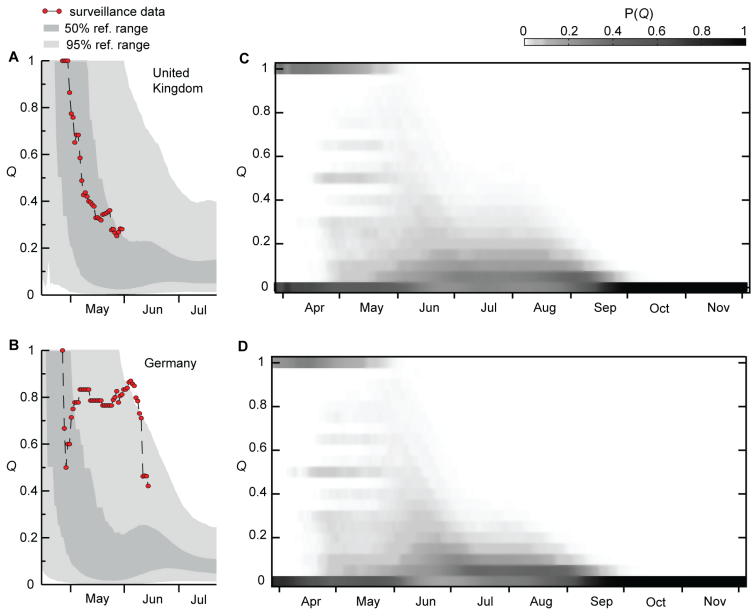

Figure 5.10: Importation of cases. A,B, Simulation results of the fraction Q of imported cases in United Kingdom (A) and Germany (B). The quantity Q is a measure of the relative weight of case importation with respect to local transmission events. The gray shaded areas show the 95% and 50% reference ranges of the simulation results obtained from 2,000 stochastic realizations. The surveillance data are indicated by red dots. While the numerical simulations are in very good agreement with the UK data, larger fluctuations are observed in Germany. It is worth to notice that the empirical data refers to confirmed cases and thus depends on the potential biases and inaccuracies of the surveillance system. Nonetheless, GLEaM simulations are still able to capture the timing of the observed drop and the values are mostly within the 95% reference range. C,D, Time evolution from April to November 2009 in the United Kingdom (C) and Germany (D) of the probability distribution to observe in any given realization of the epidemic the ratio Q between imported cases and the total number of cases. The probability distribution is reconstructed through the simulation of 2,000 stochastic realizations. Large values for the quantity Q are observed with high probability only in the early phase of the respective country's epidemic. The observed non-zero probability for a fraction of imported cases equal to zero at the early stage is due to the fact that the epidemic is imported in some cases by non-detectable individuals, such as latent and asymptomatic infectious individuals.

Adopted measure	Country	Time period
Mexican flight ban	Argentina China Cuba Peru	from April 28 to May 14, 2009 (219; 220) from May 2, 2009 (221; 222) from April 30 to May 31, 2009 (223) from April 29 to May 13, 2009 (224; 225)
Quarantine of passengers	China, Hong Kong Japan, Taiwan, Singapore.	discontinued by all countries by July 2009.
Thermal screening	Bulgaria, Chile, China, Ecuador, Hong Kong, India, Jordan, Lebanon, Malaysia, Qatar, Singapore, Thailand, UAE.	discontinued by all countries by January 2010.
Health travel warnings	Bosnia, Bulgaria, Canada, Chile, Colombia, France, Germany, Korea, Russia, Turkey, United States, UK Venezuela, Vietnam.	discontinued by all countries by June 2009.

Table 5.7: Known measures adopted worldwide against the pandemic spread (211; 226; 227; 228).

we take into account the empirically observed drop in air traffic, following the data reported in (230). The best estimates presented in section 5.2 were achieved when the informations about the reduction of passengers fluxes were not available (and not even observed). Afterwards we performed a new Monte Carlo likelihood estimate including the observed travel drop and we recovered the same results. This is mainly due to the fact that, as we will show in this section, a travel flux reduction of 40% leads to a very small delay in the epidemic propagation (3 days on average), and because the set of countries considered for the likelihood estimate were mostly infected before such reduction. The reference scenario is then compared to a set of hypothetical scenarios in which increasingly larger restrictions in individual mobility are considered, as well as different starting dates for the implementation of such restrictions. The efficacy of travel-related measures is therefore measured on the timing of seeding events and resulting delays.

Compartmentalization permits tracking of the arrival of detectable (i.e. symptomatic) and non-detectable (i.e. latent or asymptomatic) infected individuals in a given country. By defining the arrival time as the date the first symptomatic case arrives in the country under study, it is possible to quantify the delay in the spreading of the epidemic.

Figure 5.11 shows changes induced by travel restrictions on the simulated chronology with respect to the reference case by tracking the arrival time probability distribution. Results are reported in panels A, B of Figure 5.11, where application of the interventions is shown to reduce the probability values right after the peak of the distribution, with almost no change in the date of the peak. If we focus on the first arrival from Mexico, considering all possible seeding events (i.e. latent, asymptomatic, and symptomatic), we observe similar reductions in the rate of increase in the cumulative probability distribution of the seeding event, pointing to a slower rate of importation (see Figure 5.11C, D). However, the resulting change is not able to halt the spread.

By considering the time at which the cumulative probability for the seeding from Mexico has reached 90%, we can calculate the delay induced by larger reductions in air travel. It is quite impressive to notice that, according to our model, the 40% drop in travel flows observed in reality only led to an average delay in the arrival of the infection in other countries (i.e. the first imported case) of less than 3 days. We then test whether an additional decrease in travel flows of magnitudes larger than the observed 40% would have provided an additional benefit in slowing down the propagation of the H1N1 virus across the world. We consider drops in the air travel flows connecting Mexico with the rest of the world starting on April 25 following the international alert, optimistically assuming a prompt implementation by authorities with no further delays.

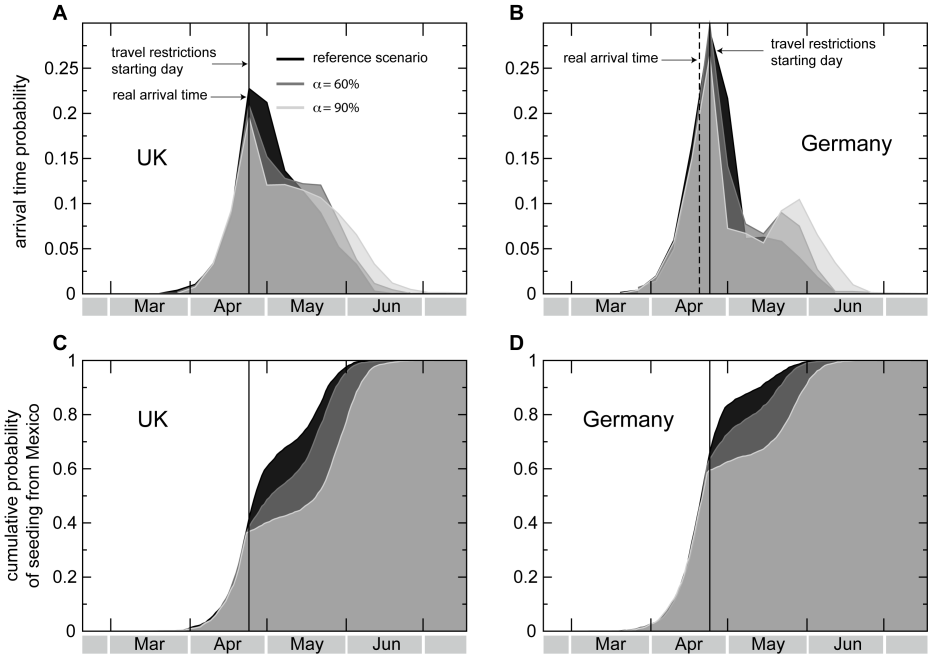

Figure 5.11: Effects of restrictions in the air travel to/from Mexico on the probability distributions of the seeding events. Travel measures imposing a reduction of $\alpha = 60\%$ and $\alpha = 90\%$ are compared to the reference scenario where the observed drop in air travel to/from Mexico is taken into account. A,B, Probability distributions of the arrival time (defined as the date of arrival of the first symptomatic case) in the United Kingdom (A) and Germany (B) for different values of α. Here we consider the importation from any possible source country, not only Mexico. The vertical dotted line indicates the observed arrival time in the country, as obtained from official reports, and the vertical solid line indicates the starting date of the travel restrictions, April 25, 2009, the day after the international alert. The probability distributions are obtained from 2,000 stochastic realizations and data are binned over 7 days. Even when imposing $\alpha = 90\%$, the peak of the probability distribution is not delayed with respect to the real scenario. C,D, Cumulative probability distributions of the first seeding event from Mexico to the United Kingdom (C) and Germany (D) for different values of α. Here we consider any source of infection in the seeding event, including symptomatic cases and non-detectable infected cases, such as latent and asymptomatic, as allowed by the computational approach. The distributions are computed over 2,000 stochastic realizations. The effect of travel restrictions is very limited in delaying the time at which the cumulative distribution reaches unity.

We also assume that the reduction is kept constant across time, differently from the empirically observed decline that successively decreased to become negligible in about 3 months.

Figure 5.12A shows the delays obtained for a selection of countries. Even given the unlikely assumption of a 90% travel reduction, the resulting delay would be of the order of 2 weeks, confirming results from previous studies (111; 124; 214; 215). This time could be used to finalize the response by the public health infrastructure of unaffected countries following the international alert, thus gaining time to enhance surveillance systems and allocate resources. Unfortunately, this timescale is insufficient to develop and distribute a vaccine. Anticipation of travel reductions following local epidemiological alerts in Mexico or the onset of symptoms from the first case in the US would have led to similar results (see panel B of Figure 5.12).

The exponential increase of cases in the outbreak region explains the negligible impact of travel restrictions over the course of the pandemic. Given two coupled populations with deterministic infection dynamics, the delay is a logarithmic function of the applied travel reduction of magnitude α, $\Delta t = -\tau \ln(1-\alpha)$, where τ is the timescale of the epidemic's exponential growth in the seed population (231; 232). The exponential increase of cases in the outbreak region is therefore responsible for the relatively limited delay induced by strong and lasting travel reductions. When $\alpha = 65\%$, $\alpha = 80\%$ or $\alpha = 95\%$ the corresponding delays become approximately 1, 1.6, and 3 times, respectively, the timescale τ that is typically on the order of a few days. The logarithmic relation also explains more realistic situations in which the epidemic origin is characterized by spatial heterogeneity and intra-region mobility that is not subject to travel restrictions. This was the case of the H1N1 pandemic, which initially diffused within Mexico before reaching international hubs and propagating internationally.

Beyond the assessment of the invasion delay, the fundamental question concerns the extent to which mobility restrictions are able to achieve containment at the source of the pandemic, especially in combination with timely mitigation policies in the country of origin. To this end we rely on a simplified modeling framework based on a metapopulation scheme describing a network of subpopulations (nodes) coupled with mobility processes (links, see Figure 5.14A) whose features reproduce the topological and mobility properties of real-world transportation systems (82; 233).

We consider a synthetic metapopulation system whose demographic and mobility properties are set in order to reproduce the statistical properties of the real systems. As briefly reviewed in chapter 2, several mobility networks at different scales – intra-city (43; 234), inter-city (14; 235), country scale (235), worldwide scale (15; 236) – and of different type – air travel (15; 236),

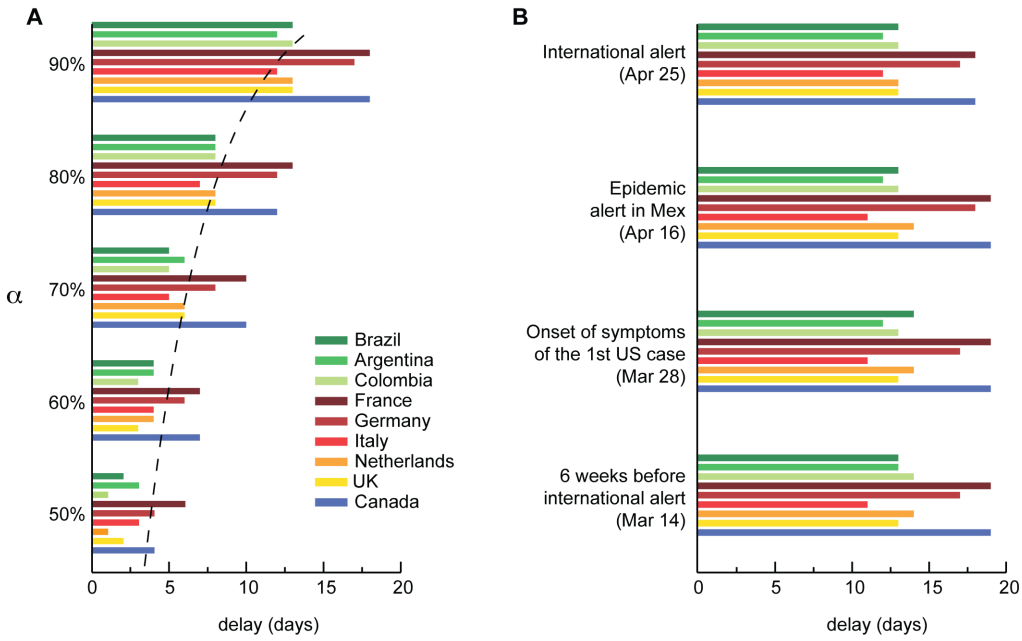

Figure 5.12: Delaying effects in the international spread. A, Delay in the case importation from Mexico to a given country compared with the reference scenario as a function of the travel reduction α. The delay is measured in terms of the date at which the cumulative distribution of the seeding from Mexico (see Figure 5.10) reaches 90%. The dotted line shows the logarithmic behavior relating the delay as a function of the imposed restrictions. The largest delay, gained when imposing $\alpha = 90\%$, is less than 20 days for all countries. The model also considers the implementation of sanitary interventions in Mexico during the early stage that was able to damp the exponential increase of cases in the outbreak zone. Travel restrictions would therefore lead to a larger impact during this phase due to the mitigating effect on the local epidemic. If a country would seeded during this phase, the resulting delay induced by the travel restrictions would have been larger, thus creating the observed differences in the resulting delays by country. B, as in A, where earlier dates for the start of the intervention are considered, at fixed $\alpha = 90\%$: April 25, corresponding to the day after the international alert; April 16, corresponding to the epidemic alert in Mexico; March 28, corresponding to the onset of symptoms of the first case in the US; and 6 weeks before the international alert. In all these scenarios and for different countries, the delay is always less than 20 days, highlighting that even the enforcement of strong travel reduction as early as possible would have had little effect.

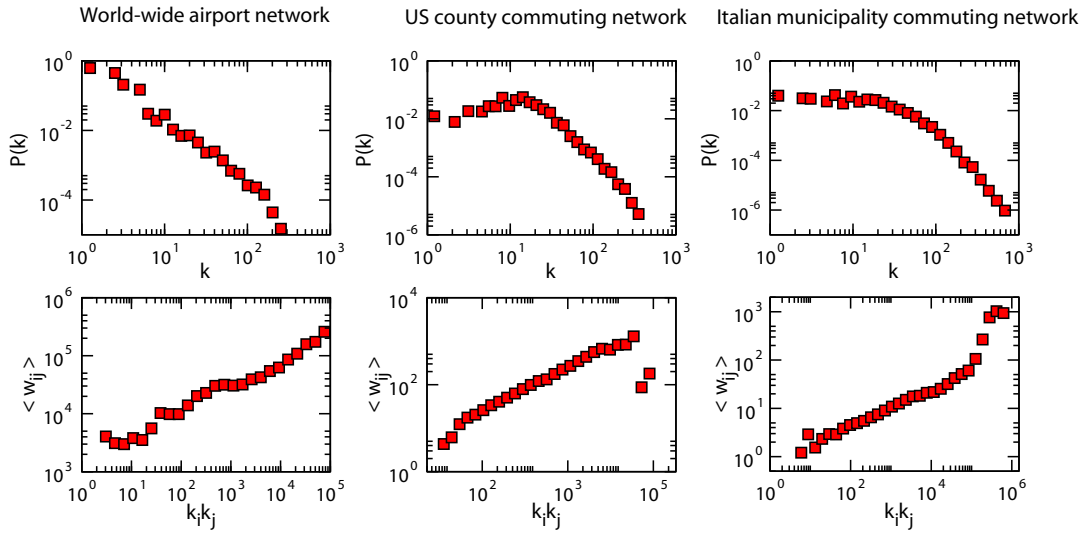

Figure 5.13: Degree distributions and average weight of the connections as a function of the product of connected node degrees for three empirical mobility networks.

commuting (14; 235), movement of people between city locations (43; 234) – have been studied and found to exhibit large-scale heterogeneities at different levels. In particular, the number of connections k from a given location is generally described by a broad distribution $P(k)$, with $P(k)$ representing the probability that a randomly extracted node has degree k. In addition, the fluxes of traveling people (the weight w_{ij} of the link connecting i to j (15)) are also found to be characterized by very large fluctuations with a weight probability distribution $P(w)$ spanning several orders of magnitude. Finally, a statistical law relating the travel flux w_{ij} to the number of connections departing from the two ending nodes i and j was found in the worldwide air transportation network (15):

$$w_{ij} \sim (k_i k_j)^{\theta}. \tag{5.3}$$

These properties are illustrated in Figure 5.13 for the case of three empirical mobility networks characterized by different spatial scales: the air transportation network analyzed in (15), the commuting patterns among counties in the United States (208) and among municipalities in Italy (237). The figure reports for the three datasets the results for the degree distribution $P(k)$, and the travel fluxes w_{ij} as functions of the topology expressed in terms of $k_i k_j$. All networks display large heterogeneities in the degree distribution and exhibit travel fluxes consistent with

Eq. 5.3. It is worth to note that these statistical features are invariant under changes of the mean of transportation and of the spatial scale, thus pointing out their robustness as peculiar aspects characterizing these systems. Disregarding the high-resolution details of numerical approaches, this synthetic metapopulation model can be analyzed, defining a new theoretical framework that allows for the study of epidemic containment measures. Starting from a single subpopulation infected at time $t = 0$, it is possible to describe the invasion dynamics at the subpopulation level in a Levins-type approach by considering the microscopic dynamics of infection and of individual travel (233). The system is characterized by a subpopulation reproductive number R_*. Analogous to the reproductive number R_0 at the individual level, R_* indicates a threshold behavior of the system: if $R_* > 1$ the epidemic reaches global invasion; otherwise, it is contained in the subpopulation of the first infected case. It is possible to derive an expression for the global invasion threshold in a branching process approximation (238; 239).

Following the empirical analysis of Figure 5.13, we assume a metapopulation model whose underlying structure is heterogeneous to include degree fluctuations, and characterized by travel fluxes following Eq. 5.3. Both topology and travel fluxes are therefore expressed in terms of the degree k of each subpopulation. A convenient description is then provided by the degree-block variables of the metapopulation system (82), where each quantity that depends on a subpopulation i (e.g. population size, number of infectious, etc.) depends only on the subpopulation degree k_i. This corresponds to a mean-field assumption for which subpopulations with a given degree k are considered statistically equivalent. The method is general and can be used for different disease compartmentalization, but here we have in mind an SIR-like model. Under the assumption that subpopulations having the same number k of connections are equivalent, we define D_k^0 as the number of diseased subpopulations of degree k at generation 0 (i.e. at the beginning of the branching process). During the entire duration of the outbreak experienced by the D_k^0 subpopulations, each of them can in principle seed some of the neighboring subpopulations thus leading to a number D_k^1 of diseased subpopulations of degree k at generation 1, for various values of the degree k. By iterating the seeding events, it is possible to describe the evolution of the number D_k^n of diseased subpopulations with degree k at generation n, yielding (233; 240):

$$D_k^n = \sum_{k'} D_{k'}^{n-1}(k'-1)P(k|k')(1-R_0^{-\lambda_{kk'}})\left(1-\sum_{m=0}^{n-1}\frac{D_k^m}{V_k}\right) \tag{5.4}$$

The r.h.s. of the equation 5.4 describes the contribution of the subpopulations of degree k' at generation $n-1$ to the infection of subpopulations with degree k at generation n. Each of the D_k^{n-1} has $(k'-1)$ possible connections along which the infection can proceed (the -1 term

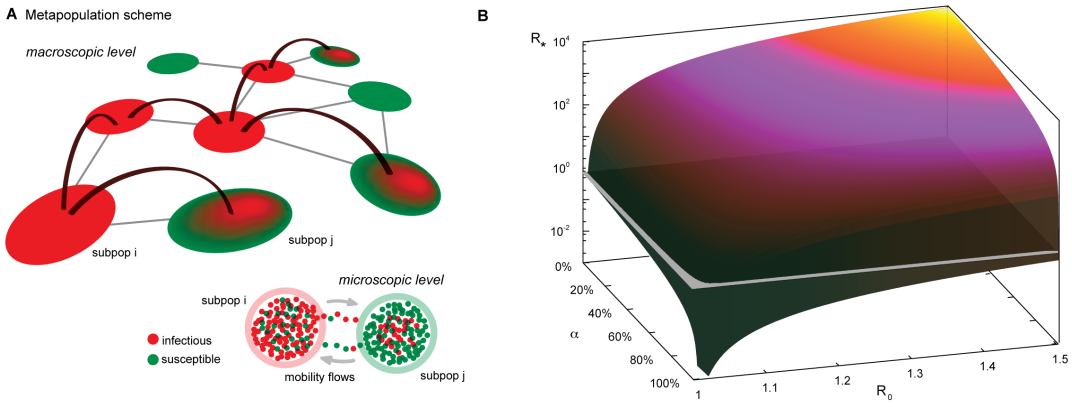

Figure 5.14: Network heterogeneity and failure of travel restrictions aimed at containment. A, Schematic illustration of the simplified modeling framework based on a metapopulation scheme. At the macroscopic level the system is composed of a heterogeneous network of subpopulations. At the microscopic level, each subpopulation contains a population of individuals. The infection dynamics are described by a simple compartmentalization (compartments are indicated by different colored dots in the picture). Within each subpopulation, individuals are mixed homogeneously and can migrate from one subpopulation to another following the mobility connections of the network. In this way the disease can spread at the subpopulations level. B, Plot of the global invasion threshold R_* described by Eq. 5.10. Here, R_* is plotted as a function of the basic reproductive number R_0 and the traffic reduction α, which is the parameter representing the percentage of variation in the total traffic w_0 in Eq. 5.10. Only in the case of extremely low values of R_0 or extremely large values of α is it possible to reduce R_* below the threshold.

corresponds to the link through which each of those subpopulations received the infection). In order to infect a subpopulation of degree k, three conditions are needed: (i) the connections departing from nodes with degree k' point to subpopulations of degree k, as indicated by the conditional probability $P(k|k')$; (ii) the reached subpopulations are not yet infected, as indicated by the term $\left(1-\sum_{m=0}^{n-1}\frac{D_k^m}{V_k}\right)$, where V_k is the total number of subpopulations with degree k; (iii) the outbreak seeded by $\lambda_{k'k}$ infectious individuals takes place, and the probability for this event to occur is given by $(1-R_0^{-\lambda_{kk'}})$ (241). The link between the microscopic dynamics of the infection transmission among individuals and the coarse grained description at the metapopulation level is encoded in the term $\lambda_{k'k}$. It represents the number of infectious people traveling from the diseased subpopulation k to the neighboring subpopulation k' during the entire duration of the outbreak, and it depends on the details of the diffusion process of individuals as well as the individual travel behavior and its interplay with the disease stages.

We assume that the rate of diffusion on any given edge from a subpopulation of degree k to a subpopulation of degree k' scales linearly with the travel flux $w_{kk'}$ from k to k' and is inversely proportional to the population size N_k of the origin location, i.e. $d_{kk'} = \frac{w_{kk'}}{N_k} = \frac{w_0(kk')^\theta}{N_k}$, where we used the statistical law of Eq. 5.3 observed in real mobility networks as shown in Figure 5.13. The term $\lambda_{k'k}$ can be explicitly written as $\lambda_{k'k} = d_{k'k}\mathcal{F}(\mathcal{D})$ where $\mathcal{F}(\mathcal{D})$ is a function of the set of disease parameters $\mathcal{D}$ that represents the number of individuals that are potential seeders during the whole unfolding of the epidemic and depends on the choice of the compartmentalization of the epidemic model. At the first stage of the infection when most of the subpopulations are susceptible $\sum_{m=0}^{n-1} \frac{D_k^m}{V_k} \sim 0$ and assuming an uncorrelated network so that $P(k|k') = kP(k)/\langle k\rangle$ (242), and a disease with a reproductive ratio close to the epidemic threshold, i.e. $R_0 - 1 << 1$, equation 5.4 can be rewritten as:

$$D_k^n = \sum_{k'} D_{k'}^{n-1}(k'-1)kP(k)/\langle k\rangle(R_0-1)w_0(kk')^\theta \cdot \mathcal{F}(\mathcal{D}) \tag{5.5}$$

Introducing the new variable $\Theta_n = \sum_k D_k^n(k-1)k^\theta$, one can re-sum the previous relation as:

$$\Theta_n = \Theta_{n-1}(R_0-1)\omega_0 \sum_k \frac{k^2-k}{\langle k\rangle} k^{2\theta}P(k)\cdot\mathcal{F}(\mathcal{D}) \tag{5.6}$$

and finally:

$$\Theta_n = \Theta_{n-1}(R_0-1)\omega_0 \frac{\langle k^{2\theta+2}\rangle - \langle k^{2\theta+1}\rangle}{\langle k\rangle}\cdot\mathcal{F}(\mathcal{D}) \tag{5.7}$$

The global invasion leading to a growing number of diseased subpopulations will occur if:

$$R_\star = (R_0-1)\omega_0 \frac{\langle k^{2\theta+2}\rangle - \langle k^{2\theta+1}\rangle}{\langle k\rangle}\cdot\mathcal{F}(\mathcal{D}) > 1 \tag{5.8}$$

In order to explicitly compute $\lambda_{kk'}$ and hence $\mathcal{F}(\mathcal{D})$ we need to specify the compartmentalization chosen for the disease modeling. Extending the analysis of Ref. (233; 240), we explore more structured and realistic compartmentalizations that take into account the presence of latent and asymptomatic individuals and envision a possible modification of the traveling behavior after presenting clinical symptoms. More in detail, considering the full compartmental model (see Figure 4.2), the number of seeds $\lambda_{kk'}$ can be approximated to the first order by

$$\begin{aligned}\lambda_{kk'} &= d_{kk'}\left[(p_t(1-p_a)+p_a)\,(\epsilon^{-1}+\mu^{-1})\mathcal{S}_\infty N_k + (1-p_t)(1-p_a)\epsilon^{-1}\mathcal{S}_\infty N_k\right] \\ &= d_{kk'}\mathcal{S}_\infty N_k\left(\epsilon^{-1} + (p_t(1-p_a)+p_a)\mu^{-1}\right),\end{aligned} \tag{5.9}$$

since each of the $\mathcal{S}_\infty N_k$ infectious individuals (with $\mathcal{S}_\infty$ being the fraction of the population that contracted that disease, namely the epidemic size (71)) can travel with rate $d_{kk'}$ during

a time period that is determined by his stage of disease. Asymptomatic individuals and a fraction p_t of the symptomatic can diffuse out of the diseased subpopulation during a time window that equals the sum $(\epsilon^{-1} + \mu^{-1})$ of the average latency and infectious periods, whereas the $(1 - p_t)(1 - p_a)\mathcal{S}_\infty N_k$ non-traveling symptomatic individuals can only diffuse during their latency state of duration ϵ^{-1}.

By explicitly introducing the expression of the epidemic size $\mathcal{S}_\infty$ for an $SEIR$ local dynamics with R_0 close to 1 (241), we obtain the following expression for the global invasion threshold R_*:

$$R_* = \frac{2(R_0 - 1)^2}{R_0^2} \left[\epsilon^{-1} + \mu^{-1} \left(p_t(1 - p_a) + p_a\right)\right] w_0 \frac{\langle k^{2+2\theta}\rangle - \langle k^{1+2\theta}\rangle}{\langle k\rangle}. \tag{5.10}$$

The quantity R_* is thus the product of three functions that depend on the disease parameters, as well as the topology and fluxes of the mobility of individuals. Travel-related interventions can be modeled as the reduction of the mobility scale w_0 or the reduction of the traveling probability p_t of symptomatic cases. The effect of such interventions is however damped by the topological heterogeneities encoded in $P(k)$ that lead to very large values of the ratio $\langle k^{2+2\theta}\rangle/\langle k\rangle$. Therefore, a reasonable reduction in the travel flows w_0 is not sufficient to decrease R_*, keeping the value well above the threshold as shown by the 3D plot reported in Figure 5.14B.

In order to better understand the crucial role of the topological heterogeneity of the mobility network, we compute R_* for a homogeneous network with the same average values of degree $\langle k\rangle$ and weight $\langle w\rangle$ as the heterogeneous one. In this case, all nodes have the same degree $\langle k\rangle$ and all the links are characterized by the same weight $\langle w\rangle$, which leads to a traveling rate $d_{kk'}$ that is simply $\langle w\rangle/N$ through all the connections. Then, the number of seeds is given by $\lambda_{kk'} = \langle w\rangle\mathcal{S}_\infty \left(\epsilon^{-1} + (p_t(1 - p_a) + p_a)\mu^{-1}\right)$. Replacing this term in Eq.5.4, we obtain

$$R_* = \frac{2(R_0 - 1)^2}{R_0^2} \langle w\rangle \left(\langle k\rangle - 1\right) \left[\epsilon^{-1} + \mu^{-1} \left(p_t(1 - p_a) + p_a\right)\right]. \tag{5.11}$$

Figure 5.15 compares the heterogeneous and homogeneous network, and shows for both cases the two-dimensional projection of the functional $R_*(R_0, \alpha)$ (α indicates the travel reduction affecting w_0 in Eq. 5.10 and $\langle w\rangle$ in Eq. 5.11, respectively). In both cases the epidemiological parameters, ϵ, μ, p_a and p_t, are set as described in section 5.2. The red and green curves indicate the epidemic threshold $R_*(R_0, \alpha) = 1$ for heterogeneous and homogeneous networks, respectively. The picture highlights how the heterogeneity of the mobility network is responsible for favoring the epidemic invasion.

Similar conclusions apply for entry screening at the airports modeled by a reduction in the traveling probability p_t, and the modeling of effective containment policies, reducing R_0 and the

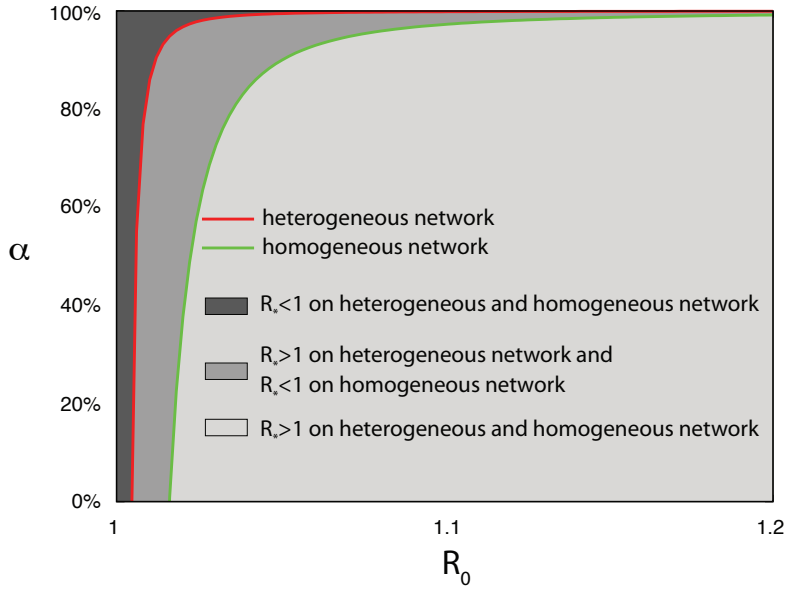

Figure 5.15: Two-dimensional projection of the functional $R_*(R_0, \alpha)$. Different gradations of grey distinguish the regions of the parameters space above and below the global epidemic threshold, while the red and the green curves indicate the epidemic threshold $R_*(R_0, \alpha) = 1$ for the heterogeneous an homogeneous network respectively.

total number of cases. The large heterogeneity of human mobility patterns is therefore responsible for the fact that travel restrictions are largely ineffective for containing an emerging pandemic.

Our analysis of the 2009 H1N1 pandemic shows that the observed decline in air travel to/from Mexico was of too small a magnitude to impact the international spread. Stricter regimes of travel reduction would have led to delays on the order of two weeks even in the optimistic case of early intervention. It is unlikely that given the ever-increasing mobility of people travel restrictions could be used effectively in a future pandemic event.

5.6 Assessment of model predictions and discussion

In the previous sections we have described GLEaM, a global stochastic simulation model of influenza epidemic based on real data on human population distribution and mobility, and its application to the 2009 A/H1N1 pandemic. We have presented the work conducted and realized in real time during the pandemic emergency that led to the publication in summer 2009 of the predicted timing for the pandemic wave in the countries of the Northern Hemisphere for the fall/winter period. In the previous section we already presented a comparison between the numerical results achieved with GLEaM and the real data regarding the contribution of imported cases to a local outbreak in the early phase of the epidemic (see Figure 5.10). In this section we present a wider comparison that considers the complete unfolding of the epidemic activity. To compare the simulated results to the observed temporal and geographic pattern of the pandemic fall/winter wave, we collected data from the surveillance monitoring systems of 46 countries around the world, accessing their official websites on a regular weekly basis and downloading their reports with the most relevant influenza activity indicators. Each surveillance system tracks different influenza activity indicators. Our data sources reported weekly at least one or more of the following indicators: ILI (Influenza-Like Illness) incidence, ARI (Acute Respiratory Infection) incidence, fraction of ILI visits or fraction of ILI patients per sentinel doctor, number of H1N1pdm laboratory confirmed cases. The indicators are generally based on the number of individuals that seek medical consultation and show respiratory symptoms that can be specifically diagnosed as influenza (ILI) or, with a broader set of possible causes, as acute respiratory infections (ARI). Even though every indicator is affected by biases, especially related to the differences in healthcare seeking behavior of different population groups, we assume that surveillance data provide a reliable estimate of the timing of the influenza activity peak, which is the only relevant information we were interested in at this stage of our work. For the same

reason, we do not address the problem of estimating the real incidence of the disease, accounting for normalization factors or different consultation rates. Since we are considering the fall wave of the pandemic only, it seems reasonable to assume a constant surveillance and consultation rate across time in a given country. The same assumption would not be true when comparing the first and the second waves, as e.g. in the United Kingdom (243). Finally, to take into account the uncertainty related to the different surveillance reporting systems, we display the observed peak weeks as a color gradient, whose limits correspond to the time interval where an incidence higher than 80% of the maximum was observed. In the Northern Hemisphere, most of the countries experienced a single major pandemic wave during autumn. The influenza activity peaked, during the October-December period, that is much earlier than the usual timing of seasonal influenza, generally peaking between January and March.

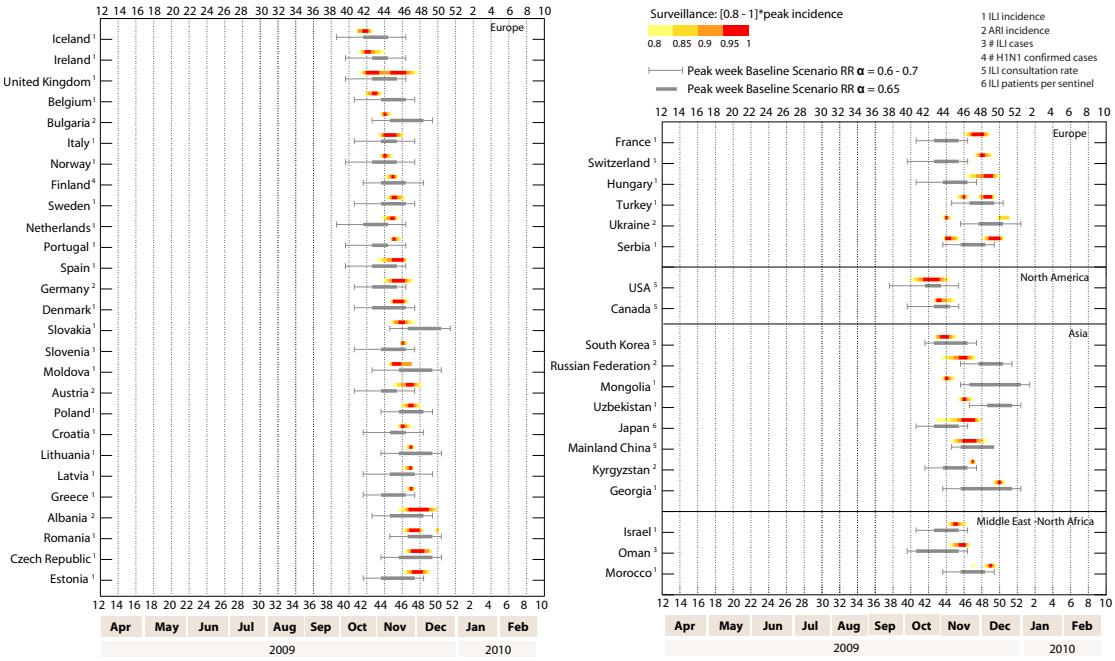

Figure 5.16: Peak timing in the Northern Hemisphere: simulations and real data. Peak weeks of the epidemic activity in the baseline scenario (grey). The reference ranges of the simulated peak week are obtained by the analysis of 2,000 stochastic realizations of the model for three different values of the seasonal rescaling factor, $\alpha_{min} = 0.6$, 0.65 and 0.7. The peak weeks reported by the surveillance are shown as color gradients, whose limits correspond to the time interval where an incidence higher than 80% of the maximum incidence was observed. Numbers from 1 to 5 indicate the kind of data provided by the surveillance of each country. Numbered weeks of the year correspond to the calendar used by the US Center for Diseases Control and Prevention.

Findings shown in Figure 5.16 indicate a good agreement in the predicted timing for a large variety of countries, including underdeveloped ones, as they all lie within the 95% reference range of the simulations, thus signaling the potentially critical importance of such an approach during an epidemic emergency.

Beyond the comparison with the actual peak times of the epidemic, further research aimed at a broader validation of the GLEaM model predictions against the data from the 2009 pandemic is still under study. As already mentioned in the previous sections, the comparison of the predicted attack rates with real data is hampered by the limited availability of accurate data on the total number of infected people (244). Surveillance data usually rely on people who seek medical care for illness, while asymptomatic individuals or individuals with mild symptoms are not counted and this leads to an underestimation of the actual number of infections. By adjusting for consultation rates, current estimates of the epidemic size range from 1.8% for symptomatic cases in the UK (243), to 18% in France for the overall proportion of infected population (245), to about $14 - 29\%$ of illness attack rate in the US (Center for Disease Control and Prevention, available at `http://www.cdc.gov/h1n1flu/estimates_2009_h1n1.htm`). The large variability of these estimates is related to the intrinsic under-ascertainment of surveillance systems, and also to different healthcare-seeking behaviors that may vary from country to country as well as change in time within the same population (see e.g. (243; 246; 247)).
The comparison of the model predictions with the empirical data would represent a valuable opportunity to assess the validity of our predictions also against the mobility network incorporated in the model. The aim would be to understand whether the full complexity of the real data considered in GLEaM was essential to obtain the predictions presented in the previous sections, or if a simplified version of the model would yield similar results. Such studies, together with the assessment of the model results against the overall burden of the epidemic, are still a work in progress and they are a matter of interest of future investigations.

6

Dynamical network analysis and spreading simulations

Contents

As reviewed in chapter 2, many empirical datasets have been analyzed by leveraging on statistical tools gathered from network science. In the last decade the development of such techniques has been refined and nowadays a taylored analysis of a networked system is able to unveil many non-trivial features and emerging phenomena. However most of the mathematical and statistical tools developed so far for the study of complex networks have been devised with the aim of understanding static topologies. In fact the lack of longitudinal data led quite often to an aggregated view of intrinsically dynamic systems as, for instance, for the pioneer investigations on the network of human sexual contacts (57). On the other hand, when the links are present for long time periods (depending on the investigated research question) a static representation of the network topology perfectly fits the real setting as in the case of the power grid (13) or inter-urban streets (14) that are static infrastructures growing over long time scales. Nonetheless in different settings it can be necessary to go beyond a static representation because the system

dynamics itself can be intricate and worthy to be addressed in order to better understand the unfolding of dynamical processes affected by the temporal evolution of the underlying evolving system. In this perspective the emerging challenge of the study of complex networks is to include the temporal dimension in the system description.

New and richer datasets including temporal information are necessary to define novel and proper observables to investigate and characterize networked systems evolving over time. In this chapter we take steps in this direction, taking as a case study the longitudinal dataset of Italian cattle movements that reports the mobility of individual animals among farms on a daily basis. The importance of a data-driven approach for the development of new mathematical tools aimed at describing dynamical networked systems resides in the possibility of comparing the dynamical behavior of different datasets, it helps in showing the limits of current techniques and it stimulates discussions and provides hints on the tools needed for a deeper understanding. Despite their importance for the spread of zoonotic diseases, our understanding of the dynamical aspects characterizing the movements of farmed animal populations remains limited as these systems are traditionally studied as static objects and through simplified approximations. Here, the complexity and inter-relations between topology, function and dynamical nature of the system are characterized at different spatial and time resolutions, in order to uncover patterns and vulnerabilities fundamental for the definition of targeted prevention and control measures for emerging zoonotic diseases. The chapter is organized as follows. After the description of the dataset under study, we first recognize the intrinsic dynamics at the agent level and then we characterize the system dynamics in terms of successive snapshots obtained from aggregating the data on different time windows. This allows us to study the emergence and robustness of network properties across time and the role of the timescale of aggregation. We then analyze the dynamical evolution of the network at the micro level, and explore its impact on the structural backbone of the system. The data analysis is then completed by introducing a novel definition of dynamical motifs for a time dependent evolving network, able to uncover causal recurrent paths in the bovine movements. We then provide insights about the efficacy of control measures against spreading phenomena. In the last section we model an infectious disease spreading on the system and we investigate the impact of the network dynamics on the unfolding of the epidemic. In particular, we assess the role of initial conditions in generating an outbreak and classify the seeds into clusters leading to similar disease invasion pathways.We investigate the temporal stability of the clusters and put forward a novel procedure to identify specific nodes

that should be monitored as disease sentinels that can be exploited for epidemic risk assessment and inform the design of optimal surveillance systems.

The results presented here are based on the following papers:

- *Dynamical patterns of cattle trade movements* published in PLoS One in May 2011,
- *Seeds clustering and sentinel identification for disease spreading on dynamical networks* submitted.

6.1 Background

The 2001 Foot-and-Mouth disease epidemic in the UK (248) represents a paradigmatic example of how the animal movements related to the trading and marketing of livestock may offer an easy mean for rapid dissemination of zoonotic infectious diseases among animal holdings, with a spatial extent covering large geographical distances. Animal diseases may compromise livestock welfare and reduce productivity, and may in addition represent a threat to human health, since the emergence of human diseases is dominated by zoonotic pathogens (249). Disease management and control is thus very important in order to reduce such risks and prevent large economical losses (250). To correctly evaluate the possible preventive and control measures, a detailed knowledge and regulation of animal movements is needed. A crucial step into addressing this issue has been taken after the Bovine Spongiform Encephalopathy crisis of 1997, when the European Union Council imposed the systematic identification and traceability of individual bovine animals (251), and similar cattle identification and tracing systems have also been implemented in other countries (252). Large datasets describing the cattle movements at the individual animal level on a daily basis have thus become available. Such monitoring efforts have led to a unique opportunity of studying animal movements in a detailed way, characterizing their behavior in time and space, and identifying patterns that may become relevant for the spread of a potential disease in the cattle population. A natural description of these systems is offered by the network representation in terms of nodes (the elements of the system, i.e., the premises in the cattle flow case) and links (the interactions among its elements, i.e., the cattle movements among premises) (8; 27; 37; 38; 39; 253). The application of network approaches to veterinary medicine is however rather new. As reviewed recently by Dubé et al. (254) and by Martinez-López et al. (255), few papers have been published that analyze livestock movements by constructing the network of displacements and studying the relations between nodes with a systemic approach, thus going

beyond the simple characterization of single-node properties (as e.g. the amount or frequency of displacements on and off single farms). Most of the veterinary studies are indeed based on static representations of cattle flows - where the temporal information of the displacements is collapsed into few successive snapshots of the datasets (256; 257; 258; 259; 260; 261; 262; 263; 264), or explored through the time series analysis of simple global quantities (257; 258; 261; 262; 265) - or focus on the results of spreading simulations based on the dynamical network and on its static counterparts (266). In the analyses performed so far, results have shown a large heterogeneity in the connectivity patterns among premises, with probability distributions for the number of incoming and outgoing connections (in-degree and out-degree, respectively) characterized by broad tails (256; 257; 258; 260; 261). Such results are typically obtained from the investigation of a static network obtained by aggregating data on the full available time window (256; 257; 260), and few examples of structures extracted from shorter aggregation times (such as e.g. monthly and weekly networks) have been investigated (258; 261), without, however, exploring in a systematic way the stability of the observed features across time. With the aim of assessing the spreading potential induced by the complex structures hidden in the data, much work has been dedicated to the analysis of the components of the network (giant component, weakly and strongly connected components, etc.) in order to estimate the upper bounds of the epidemic size (256; 258; 260; 261; 265; 267), and to the ranking of nodes in terms of various measures of centrality defined a priori, such as degree, betweenness, and others (256; 259; 267), that in other systems were found to impact the behavior of dynamical processes taking place on top of them (37; 77; 82; 268; 269; 270; 271; 272; 273; 274; 275; 276; 277; 278). The efficacy of prevention and control measures based on this information (256; 258) have been studied, though no assessment of the stability of these features in time is provided, thus affecting the applicability of the same measures in different points in time, due to the time evolution of the network.

6.2 Data description

Data on cattle trade movements were obtained from the Italian National Bovine database, which is administered by the Italian National Animal Identification and Registration Database (279). The database details the movement of the entire Italian population of bovines among animal holdings, providing a comprehensive picture of where cattle have been kept and moved within the country. Each movement record reports the unique identifier of the animal, the codes of

the holdings of origin and destination, and the date of the movement. Such tracking system allows us to easily reconstruct the path of each bovine and to build the corresponding overall network, minimizing the problems related to data accuracy that are found in other tracking systems that do not provide both origin and destination of the displacements (260; 261). Additional information was provided for the animal holdings, including the type of premises (i.e. fattening farm, dairy farm, pasture, slaughterhouse, assembly center, market, genetic material center, and other), and their georeferenced metadata in terms of the geographic coordinates of the centroids of the municipality where the premises were located. Here we examine the records for the year 2007 (256). A total of 4,946,201 bovines were tracked, counting for 7,177,825 recorded displacements of individual animals and 1,592,332 distinct batches movements. There were 173,139 active premises during the year (i.e. they either received a batch or moved it), of which 49.9% were fattening farms, 26.1% were dairy farms, 1.7% were pasture, 1.1% were slaughterhouses, 0.4% were assembly centers, 0.06% were markets, 0.04% were genetic material centers, and the remaining 20.7% were labeled as other premises. Active premises are located on almost the entire territory of the country, covering 96% of the Italian municipalities, though their distribution is not uniform - a single municipality can indeed contain a number of holdings varying from few units to hundreds. A total of 365 days of activity was recorded, from January 1st to December 31st of 2007, signaling that at least one displacement per day took place in the year under study. The dataset also contains information on the importation and exportation of cattle; these movements, representing less than 1% of the total number of movements in the database, were however excluded from the analysis as the focus of our study is on the full set of displacements within national boundaries. Table 6.1 summarizes some basic properties of the dataset.

Differently from human mobility data where the information is usually not provided at the individual level and is aggregated into flows that cannot be traced back to the individual's behavior (15; 78), the cattle movement dataset provides detailed information at the individual level through tracking each single animal during its displacements. This allows two different levels of description of the dynamics: (i) the agent-centered point of view that considers the features of the animals' movements (similarly to what can be done for individuals based on anonymized phone cell data (21; 48)); (ii) the network point of view that focuses instead on the system's behavior and is given by the evolution of the topology, and of the links' and nodes' properties from one time window to the next. These views provide complementary information for the characterization and understanding of the dataset. Gaining insight from the agent-centered

Table 6.1: Cattle trade movements: data from the Italian National Bovine database for the year 2007.

property	value
tracked bovines	4,946,201
number of animal movements	7,177,825
number of cattle movements	1,592,332
average batch size	4.5
days of activity	365 [Jan 1- Dec 31]
municipality involved	7,780 (96% of the Italian municipalities)
number of active farms	173,139
average number of active farms per municipality	25

dynamics aims at characterizing the trajectories of each bovine, uncovering the possible presence of predictable patterns, similarities or large heterogeneities, in the perspective of understanding the potential for disease propagation across the system, through its agents.

As bovine displacements are subject to livestock commercial constraints, we expect that the resulting bovine mobility patterns will be different from human mobility patterns (21; 22; 48; 78). Indeed, the number of displacements of any single animal over one year is quite restricted (257; 280), as shown in Figure 6.1 , in particular if compared with human behavior. On average bovines experience 1.45 displacements during a year. Interestingly however, some animals perform more than 10 moves, which may potentially result in superspreader behaviors. Figure 6.1 also reports the distributions of geographical distances covered either in a single displacement (i.e., the geographical distance between the origin and destination farms), or following the trajectory of a single animal in one year. Despite a well-defined maximum at short distances, these distributions display rather broad tails corresponding to very long routes (257; 261). In addition, the distributions are robust against filtering on the yearly batch size displaced, indicating that very long routes are performed by both small and large batches. The possibility of such long displacements should be taken carefully into account when dealing with spreading of diseases, as they could result in epidemics rapidly reaching geographically very distant parts of the system.

Another interesting issue concerns the time interval between two consecutive displacements of an agent, corresponding to the period that a given bovine spends in the same holding (257). This time interval may represent the time of exposure of the animal to a potential outbreak taking place in the holding, or the time during which it could spread the disease to other animals

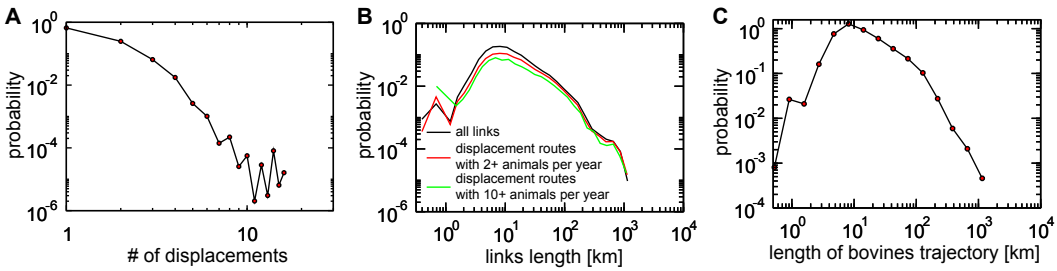

Figure 6.1: Bovine activity. Panel A shows the probability distribution of the number of displacements that a bovine experiences during one year. Panel B displays the probability distributions of the distances covered during a single displacement. Since many links correspond to the displacement of very few animals, the same distribution is shown with different thresholds, i.e. considering only routes with at least 2 or 10 bovines displaced during the year under study. This corresponds to keeping respectively 42% and 13% of the original displacement routes. Panel C shows the probability distribution of the distances covered by a single animal during its trajectory in one year.

if infected. Since the different types of farms have different roles in the bovines trade, the global distribution shown in Figure 6.2 A is a convolution of several different behaviors. In particular, the two peaks at 3 and 6 months correspond to pasture and fattening farms, respectively, as shown by the other panels of the Figure that disaggregate the results by premises type. Except for the markets, in which bovines spend only few days, the distributions of these time intervals are broad for each farm type, with different slopes. This points out the large variety of possible timescales characterizing the time during which an animal stays in a given premises, indicating that homogeneous assumptions on the length of stay of an animal at a given holding do not provide an accurate description of reality. The broadness of these distributions should therefore be taken cautiously into account in the modeling approaches.

6.3 Daily and aggregated networks

The system of cattle trade movements can be represented in terms of a network, similarly to other mobility datasets and transportation systems (14; 15; 16; 21; 22; 43; 46; 78; 281; 282; 283). The simplest representation is obtained when nodes correspond to premises, and a directed edge is drawn between two nodes whenever a displacement of bovines occurs between the corresponding premises. Since data on cattle movements is provided on a daily basis by the original dataset, it is thus possible to construct 365 daily networks, each containing the activity of nodes and links for one day. It is also useful to construct static snapshots of the system by aggregating the

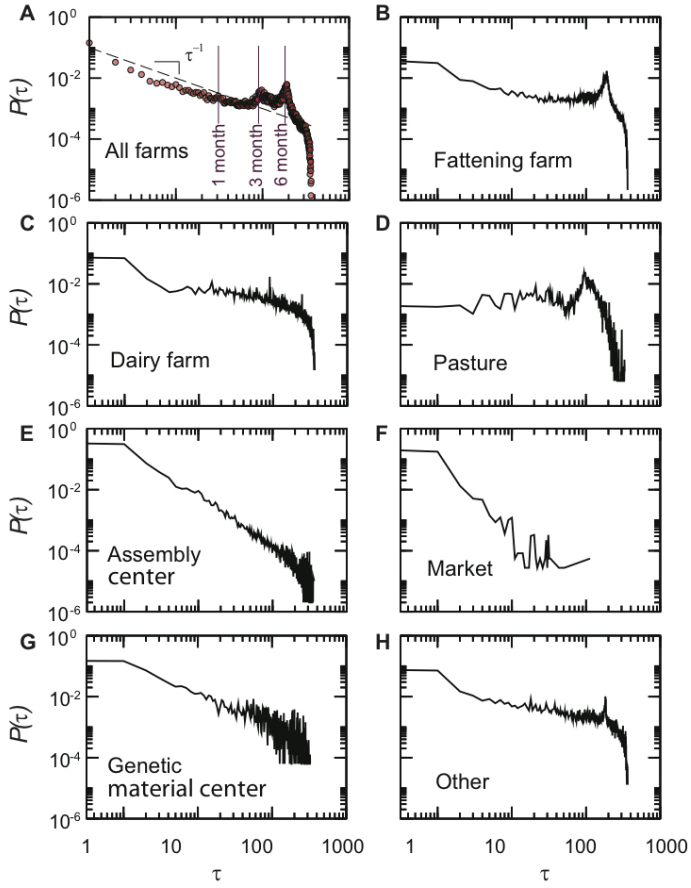

Figure 6.2: Probability distributions of the time interval τ between two consecutive displacements of a bovine. τ corresponds to the time during which the bovine stays at given premises. The seasonality behavior of breeding is clearly shown by the peaks at 3 and 6 months, while at shorter times the distribution behaves as τ^{-1}. The global distribution is a convolution of the time distributions obtained for different farm types, shown in panels B to H.

observed activity over various time windows Δt in order to assess how the choice of the time window affects the observed statistical properties and how such properties change across time for a fixed Δt. This series of static views partially looses the intrinsic dynamical nature of the system within the given time window, however it allows to study the static snapshots with the usual techniques of network theory (27; 37; 38; 39; 253; 256; 257; 258; 259; 260; 261; 262; 263; 264). Given a specific choice of Δt, we can construct $365/\Delta t$ such consecutive snapshots, corresponding to the time windows $[n\Delta t, (n+1)\Delta t]$, with n going from 0 to $365/\Delta t - 1$. In addition to the intrinsic time resolution of the system, $\Delta t = 1$ day, we also consider time windows of $\Delta t = 7$ days, $\Delta t = 28$ days (we avoid aggregating over calendar months to avoid fluctuations due to the different duration of the months during one year), and $\Delta t = 365$ days. These choices give rise to 365 daily networks, 52 weekly networks, 13 monthly networks, and one annual network, respectively, the latter aggregating the whole activity reported in the dataset. While in the literature annual and monthly networks have been typically analyzed (with the exception of Ref. (258) that studied the weekly networks of French data), here we consider different values of Δt in order to systematically explore the dynamical features of the networks on a variety of timescales. The simplest dynamical information is given by the evolution of the sizes of the aggregated networks. The numbers of nodes and links follow consistent patterns (as shown in Figure 6.3) with both weekly cycles and clear seasonal properties that distinguish the livestock activity across the different seasons (257; 258; 261; 262; 265). On a monthly time scale, it emerges that the summer activity is substantially lower than the activity registered during the rest of the year. The evolution of the daily snapshots sizes shows moreover how the overall movements decrease strongly during the weekends, leading to increasingly smaller and more fragmented networks that put obstacles to the propagation of a disease across the system.

The analysis of the various static snapshots gives access to a first characterization of the system under consideration, investigating both its structural and dynamical properties. This allows for the first time the comparison of the features obtained at different timescales, and, for each timescale, the possible emergence of properties that remain stable or change across time, as the activity captured in each snapshot may indeed vary from one snapshot to another. Even the very basic features of the network, such as e.g. the number of nodes (noted N) and of edges, depend both on the time of the year at which we observe the system and on the duration of the aggregation Δt (258). As the networks are directed, each node i is characterized by both its out-degree $k_{i,out}$ (i.e., the number of premises to which a movement is registered within the given time window) and in-degree $k_{i,in}$ (i.e., the number of premises from which the node

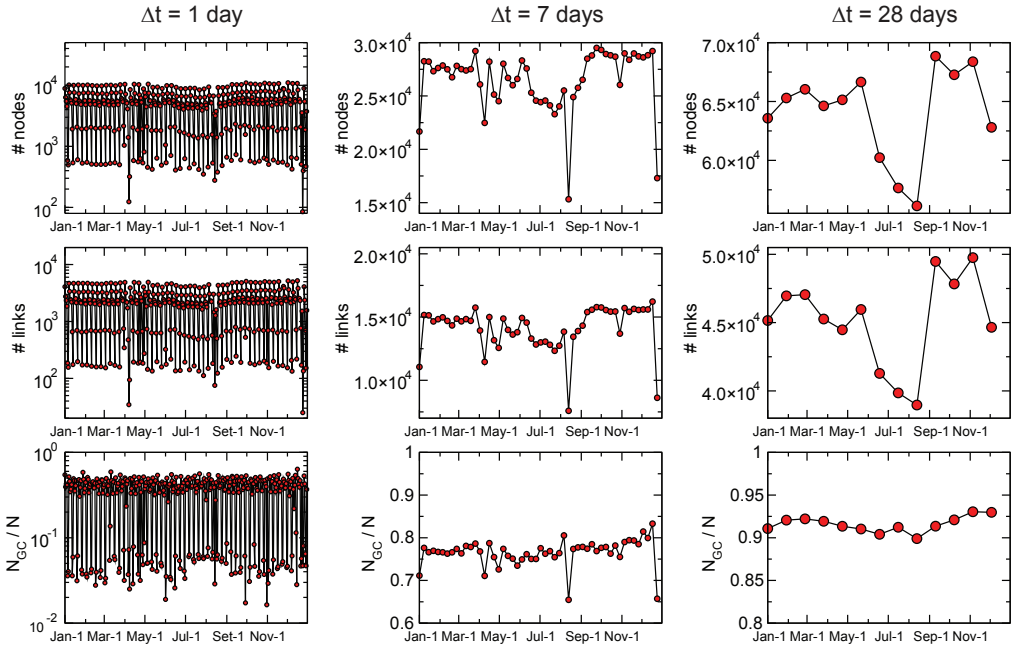

Figure 6.3: Time evolution of the global static features of networks on different timescales. The timeline of the number of nodes (top), the number of links (center), and the fraction of nodes in the giant component (bottom) are shown for daily, weekly, and monthly networks. Clear weekly and seasonal patterns are detected.

receives an incoming flux of animals within Δt). For each snapshot, we consider only nodes with $k_{i,in} + k_{i,out} > 0$, defining them as active nodes since they correspond to premises that have registered at least one incoming or outgoing displacement during the aggregation time window. Moreover, the links of these networks can be weighted according to two distinct definitions, measuring either the number of cattle batches moved or the total number of animals moved (260). More specifically, we denote by w_{ij}^B the amount of cattle batches movements recorded within the given time window Δt from the holding i to the holding j. The weight w_{ij}^A instead indicates the total number of bovines moved from i to j during Δt. The first quantity provides a binary information on a daily basis, and counts the number of movements occurring in the time window Δt under consideration; the second measures the magnitude of the movements. The introduction of two different definitions of the weight is useful in order to explore whether there are any trivial correlations among the two quantities, and to assess the limits of the approximation that uses less detailed data such as the number of movement batches, which are usually more readily available than the detailed movements of animals at the individual level (266). This would be very important in the framework of modeling approaches based on real data. By following the usual definition of strength of a node in a weighted network (15), we denote by $s_{i,in}^{B(A)} = \sum_j w_{ji}^{B(A)}$ and $s_{i,out}^{B(A)} = \sum_j w_{ij}^{B(A)}$ the in-strength and out-strength of node i, respectively, quantifying the total numbers of incoming and outgoing batch (B) and animal (A) movements of the corresponding premises during Δt. Table 6.2 summarizes the basic properties of the aggregated networks for the various values considered.

At the smallest possible aggregation scale, $\Delta t = 1$ day, the networks are small and very sparse, including an average number of nodes of the order of few thousands (to be compared to the total of about 10^5 nodes active across the whole year), and they are typically composed of small disconnected components, similarly to what was observed in the UK cattle data (280). For larger values of Δt, i.e. longer aggregation times, an increasing number of nodes and links are present in the networks, since more and more distinct displacement events are registered during the time window. The average number of nodes of the weekly and monthly networks increases of one order of magnitude with respect to the daily case, as observed in the weekly and monthly snapshots of the French cattle data (258). This number does not show great variations in time for a given Δt, however it remains very small if compared to the full network, thus indicating the presence of strong changes in the activation of nodes from one month to the other. When aggregating over time windows of increasing duration, the networks not only increase in size but also become denser, with the number of active connections growing faster than the number

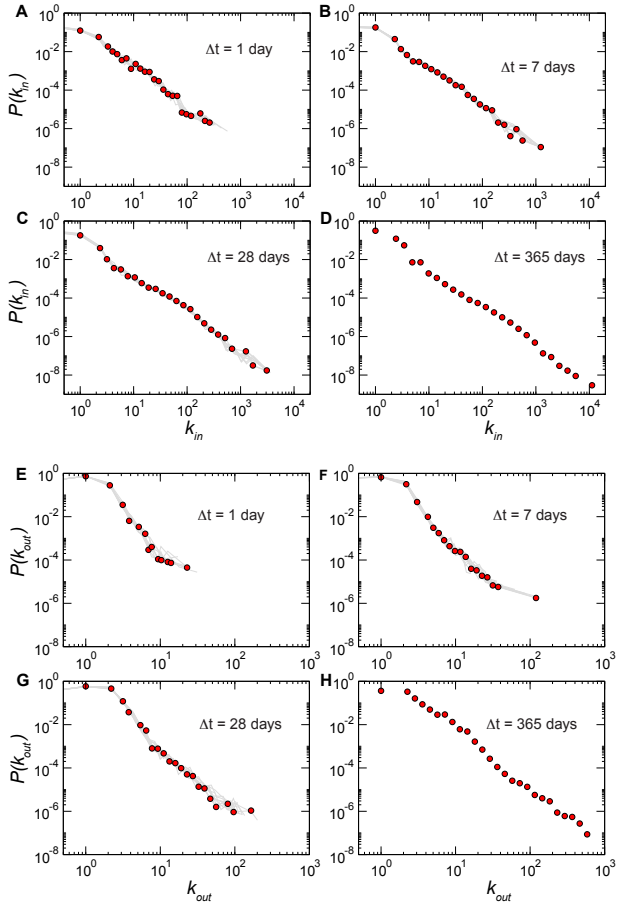

Figure 6.4: Degree distributions for networks aggregated on different timescales . Since a single value of Δt (for $\Delta t < 365$ days) yields multiple snapshots, each panel shows one distribution obtained for a given snapshot (circles) superimposed to a subset of the distributions obtained for the other snapshots at the same value of Δt (grey lines). Panels A to D report the distributions of the in-degree k_{in}, that show very large fluctuations and a power-law like behavior with exponent close to -2 in all cases. Panels E to H present the distributions of the out-degree k_{out}, characterized by a cut-off that strongly depends on the length of the aggregating time window.

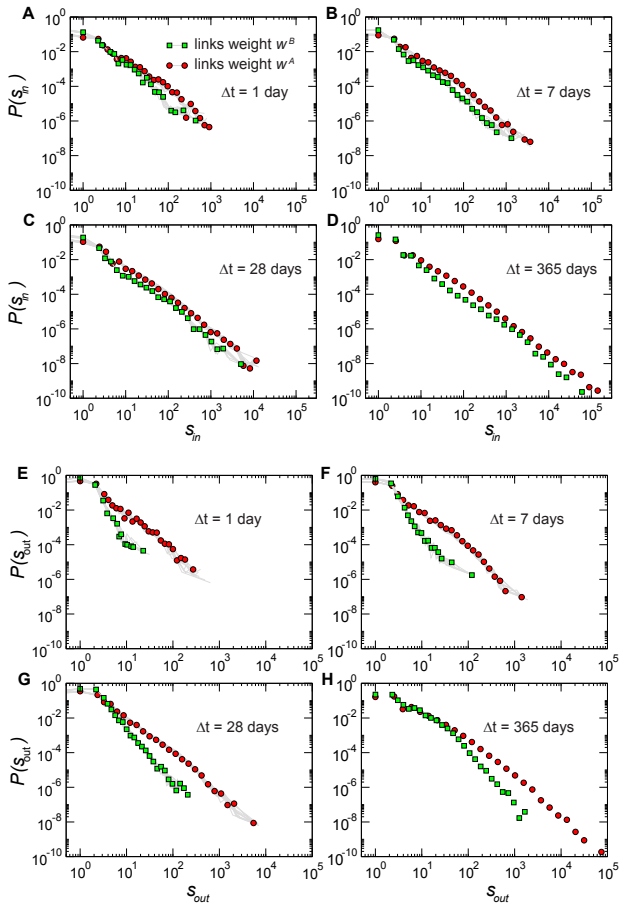

Figure 6.5: Strength distributions for networks aggregated on different timescales Δt. Panels A to D report the distributions of the in-strength s_{in}. Interestingly, the definition used to weight the links does not affect the distribution of the incoming traffic: the distributions $P(s_{in}^A)$ and $P(s_{in}^B)$ are very close. Panels E to H present the distributions of the out-strength s_{out}, whose behavior instead depends strongly on the type of weight considered. Broader tails are observed when considering the total number of animals displaced out of a given holding. The same representation of Figure 6.4 is adopted, with symbols representing the result of a particular snapshot, and grey lines the results obtained for a subset of the other snapshots.

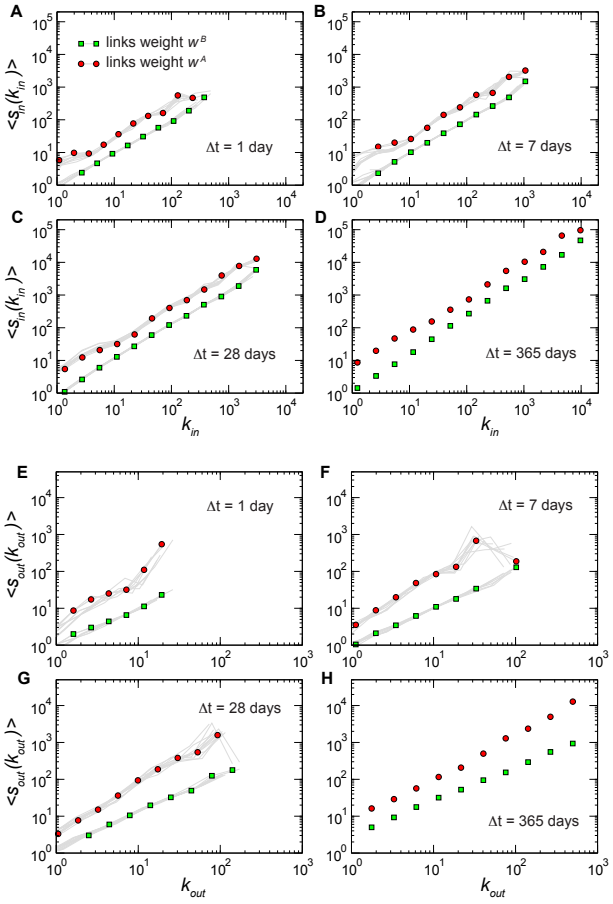

Figure 6.6: Relation between the number of bovine traffic movements of a holding and its number of connections for different values of Δt. Panels A to D report the average in-strength of nodes with a given value of in-degree, whereas panels E to H present the average out-strength of nodes with given out-degree. The same representation of Figure 6.4 is adopted, with symbols representing the result of a particular snapshot, and grey lines the results obtained for a subset of the other snapshots.

Table 6.2: Summary of the main features of the mobility networks obtained by aggregating the data over a time window Δt.

Aggregating time window	variable	average	σ	$[min, max]$
	# of nodes	$4.9 \cdot 10^3$	$3 \cdot 10^3$	$[85, 1.1 \cdot 10^4]$
$\Delta t = 1$	# of links	$4.2 \cdot 10^3$	$2.8 \cdot 10^3$	$[49, 10^4]$
(365 networks)	k_{in}	0.9	6.2	[0,683]
	k_{out}	0.9	0.8	[0,178]
	w_{ij}^A	3.6	10.4	[1,2039]
	w_{ij}^B	1	0	[1,1]
	# of nodes	$2.6 \cdot 10^4$	$2.8 \cdot 10^3$	$[1.5 \cdot 10^4, 2.9 \cdot 10^4]$
$\Delta t = 7$	# of links	$2.8 \cdot 10^4$	$3.4 \cdot 10^3$	$[1.5 \cdot 10^4, 3.2 \cdot 10^4]$
(52 networks)	k_{in}	1.1	11.9	[0,1595]
	k_{out}	1.1	1.1	[0,178]
	w_{ij}^A	3.8	11.9	[1,2039]
	w_{ij}^B	1.05	0.3	[1,7]
	# of nodes	$6.4 \cdot 10^4$	$3.8 \cdot 10^3$	$[5.6 \cdot 10^4, 6.9 \cdot 10^4]$
$\Delta t = 28$	# of links	$9 \cdot 10^4$	$6.5 \cdot 10^3$	$[7.8 \cdot 10^4, 9.9 \cdot 10^4]$
(13 networks)	k_{in}	1.4	22.9	[0,4154]
	k_{out}	1.4	1.9	[0,219]
	w_{ij}^A	4.8	18.1	[1,2039]
	w_{ij}^B	1.3	0.8	[1,25]
	# of nodes	$1.7 \cdot 10^5$	-	-
$\Delta t = 365$	# of links	$5.77 \cdot 10^5$	-	-
(1 network)	k_{in}	3.3	59.5	[0,13186]
	k_{out}	3.3	7.0	[0,649]
	w_{ij}^A	9.8	65	[1,10845]
	w_{ij}^B	2.7	5	[1,250]

of active nodes. The small disconnected components observed in the daily networks coalesce, leading to an increasingly larger giant component (i.e. the largest connected component of the network). Though snapshots up to monthly networks are small in size compared to the total number of active nodes observed during the year, their structure and interconnectivity allows for the creation of giant components spanning a large fraction of the aggregated networks , similarly to what was observed in the analysis of cattle movement data in other countries (258). Starting from daily networks that may offer only limited propagation at the daily scale, a giant component emerges if aggregating on timescales $\Delta t \geqslant 7$ that indicates the existence of paths of propagation

from one node to another at the system level.

Figures 6.4 to 6.6 report a set of statistical properties of the networks generated by aggregating the data on time windows of lengths Δt =1, 7, 28, 365 days. Given that each Δt value corresponds to a set of snapshots (except in the case of Δt = 365 days), for the sake of visualization we show in each plot the distribution of the quantity under study for one particular snapshot chosen as an example (red circles), overlaid to gray lines that indicate the behavior displayed by the other snapshots corresponding to the same Δt (in the weekly and daily cases, given the large number of snapshots, we show a random subset). This allows us to monitor the variations over time signaling changes of the system's statistical properties, as a function of changes in Δt and the time of observation. Interestingly, these distributions are superimposed for successive time snapshots at a fixed value of Δt, denoting a statistical stationarity of global distributions, describing the activity taking place at the microscopic level. This behavior, which is observed here for the first time for cattle movement data, is consistently present for all Δt under study, and is similar to what was observed in other systems for which longitudinal data is available, such as e.g. the airline transportation system analyzed in Ref. (49). Figure 6.4 displays the distributions of in- and out-degrees. The in-degree distributions are broad, with a behavior close to a power-law and a slope approximately equal to -2. This is in agreement with the results found for a specific month of the 2005 UK cattle data (261), and shows that this behavior is a common feature of the system in various countries and, moreover, is independent of Δt. The range of values of k_{in} clearly increases with increasing values of Δt. Large fluctuations are observed also in the out-degree distributions, however the range of possible values of k_{out} is systematically one order of magnitude smaller than the corresponding range observed for k_{in}, not only for the annual network (260) but for every timescale investigated. Results show a clear asymmetry in the receiving and sending activities of the animal holdings, which can be explained by the typical activity of specific premises types, such as slaughterhouses, assembly centers and also markets. Such premises are indeed responsible for assembling cattle trade fluxes for commercial purposes, thus receiving batches from a large number of premises, assembling them and moving larger fluxes to fewer premises. Similar probability distributions can be computed for the weights as well, taking into account the two possible definitions. Figure 6.7 shows how the weights w^B have by definition a sharp cutoff at their maximum value Δt, therefore limiting the range of possible values they can assume in the case of small Δt. On the other hand, the number w_{ij}^A of animals displaced between farms i and j is characterized by a broad distribution even

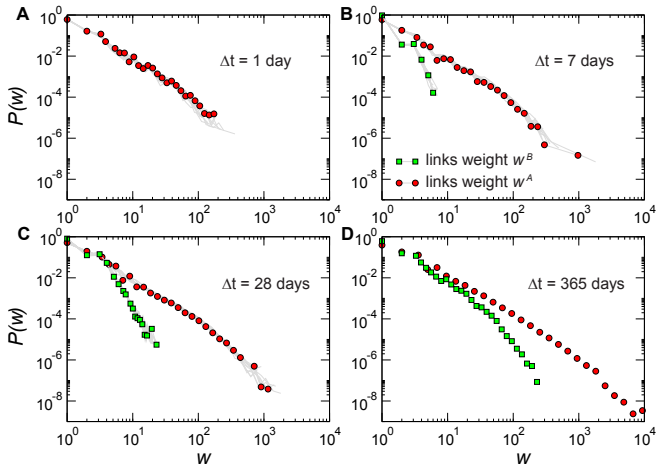

Figure 6.7: Weight distributions for networks aggregated on different timescales Δt. Red circles refer to the binned distributions of the weight w_{ij}^A, measuring the number of animals moved along the link ij, whereas green squares refer to the binned distributions of the weight w_{ij}^B that counts the number of batches displaced along the link. The same representation of Figure 6.4, with symbols representing the result of a particular snapshot, and grey lines the results obtained for a subset of the other snapshots. The cut-off of the w_{ij}^B distributions is naturally fixed by the choice of the aggregating period Δt. The distribution of w_{ij}^B for the daily networks has been omitted, since it is equal to 1 for $w_{ij}^B = 1$ and 0 elsewhere.

for the shortest time window $\Delta t = 1$ day (257). This shows how cattle displacements are most often characterized by a small number of animals, but that movements of very large numbers are also observed with a non-negligible probability. Interestingly, the shapes of the distributions are almost not affected by changes in Δt, denoting underlying non-trivial mechanisms that make these statistical properties stable across integrations on diverse timescales. Figure 6.5 shows the in- and out-strength distributions, according to the two definitions for the weights. A pattern very similar to the degree distributions is observed: the in-strength distributions are broad even at small Δt, while the out-strength distributions broaden significantly only as Δt increases, especially for s^B_{out}, which indicates the total outflow of batches. The asymmetry discussed above is thus retained if we consider the total number of animals displaced in and out of premises.

Besides looking at the overall behaviors of these quantities in terms of probability distributions, it is interesting to explore whether non-trivial correlations arise that relate the topology with the flows at the premises level, by considering the correlations between the strengths and degrees of nodes. Figure 6.6 shows the results obtained when the strengths are defined in terms of the weights w^B and w^A, considering both the inflow and outflow dynamics. The behavior is linear for the in-strength, signaling an absence of correlation between the number of premises from which a specific holding receives batches and the number of batches or bovines received on each connection (15). In the case of the out-strength we observe instead a slightly superlinear trend when s^A_{out} is expressed as a function of the out-degree, showing that more active farms in terms of number of connections also tend to send more animals on each connection (15), explaining the asymmetry observed before in the variations of s_{out} and k_{out} with respect to s_{in} and k_{in}. For increasing time window lengths, the aggregated networks take into account more displacement events. Concerning the links and nodes present in a network at a given timescale Δt, this means that their weights, degrees and strengths are expected to increase when longer time windows are considered. Notably however, we do not observe a simple shift of the whole distributions towards larger values with a corresponding absence of small values: the distributions continue to be broad, spanning several orders of magnitude, but the most probable values remain very small. In the case of the degree distributions, this can be due to nodes that have very few connections for any time window, or to nodes that are active only very rarely. For the weights distributions, it shows that on any timescale there exists many links that are active only during few days, already indicating the presence of a non-trivial underlying dynamics that cannot be uncovered through the analysis of static snapshots only.

6.4 Network microscopic dynamics

The results of the previous section show how the microscopic dynamics of cattle movements is described by statistical properties that are found to be stationary, with a behavior that is qualitatively invariant with respect to changes in the timescale (whereas size and magnitude of fluctuations clearly depend on the time window Δt). Here and in the following subsections we aim at characterizing the underlying dynamics to uncover higher order correlations and relevant temporal aspects leading to the observed behavior. By comparing the results obtained for the weekly and monthly networks with those corresponding to the whole dataset, it is clear that a strong dynamical activity shapes the evolution of the system on both global and local scales. As an example, we show in Figure 6.8 a visualization of a subgraph for three consecutive monthly networks. The subgraph is constructed by selecting a particular seed node (the same for all three networks) and by considering all nodes at distance ≤ 3 from the seed (where the distance is defined by the number of links traversed on the shortest path connecting the two nodes). Nodes keep their position in the visualization if they are active over multiple snapshots. The figure highlights how the structure of the neighborhood of a given node obtained at consecutive time snapshots can widely differ: even highly connected nodes in one snapshot can disappear from the neighborhood of the given node in the next snapshot, and hubs suddenly appear that were absent from the previous snapshot.

6.4.1 Activity timescales

Similarly to the dynamics of single animals, the network dynamics can be first characterized by the distributions of the activity and inactivity periods of nodes and links (49). These periods are defined, for a given timescale Δt, as the number of consecutive time steps in which a node, or a link, is active (or not active, respectively). In the case of time windows of $\Delta t = 1$ day, we remove the weekends from the dataset as they are characterized by a much smaller activity, and consider a node or a link to be continuously active if it is present in the snapshots of a given Friday and of the Monday of the following week. The corresponding distributions are shown in Figure 6.9 for $\Delta t = 1$ and 7 days. As seen also in the dynamics of the air transportation network (49), most nodes and links turn out to be continuously active or inactive for only very short periods. The distributions of activity periods are rather narrow in the case of daily networks, and can be fitted by power-laws with exponent smaller than -4: most nodes and links are active only for one day at a time, and only very few are continuously active for more than a few days. The

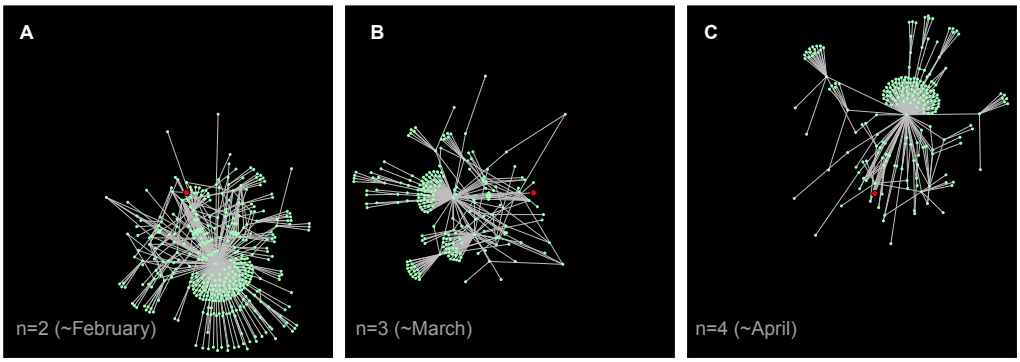

Figure 6.8: Neighborhoods of a selected node in three consecutive monthly networks. The subgraphs are obtained by showing all nodes within distance 3 from a selected node (in red in the figure), for consecutive monthly snapshots. The visualization highlights how the neighborhood of a given node may strongly change its structure in time. It is important to note that nodes that disappear from the plots may still be present in the network, but are not shown as they may be at distance larger than 3 from the seed, thus not belonging to its neighborhood.

distributions become significantly broader when considering weekly networks, where power-laws with exponents close to -3 emerge. The difference observed by comparing $\Delta t = 1$ and 7 days can be easily explained by the integration over multiple days in the case of $\Delta t = 7$: being active in two consecutive such networks is a less stringent condition than being active each day of two successive weeks. The inactivity periods $\Delta\tau$ are characterized by much broader distributions extending on all possible timescales, signaling that a node (or a link) may become active at a given point in time without then participating to the dynamics for a long time interval. From the point of view of control policies, such long inactivity periods would help in limiting the spread through self-isolation of premises.

Given that the activities of nodes and links of the displacement network occur at both short and long timescales, here we aim at characterizing the mechanisms behind the appearance and disappearance of links in the system, and we focus on the weights w^A that measure the number of animals displaced along each link. As proposed in (49) we evaluate in particular the fraction of appearing f^a and disappearing f^d links, as a function of their weight, in order to uncover a possible correlation between a link's stability and the number of displaced animals along that link. More precisely, if $E(w|t)$ is the number of links with weight $w = w^A$ at time t and $E^a(w|t)$ is the number of such links that were not active at the previous time (and thus appeared at time t), the fraction of appearing links is $f^a(w) = E^a(w|t)/E(w|t)$. An analogous procedure leads

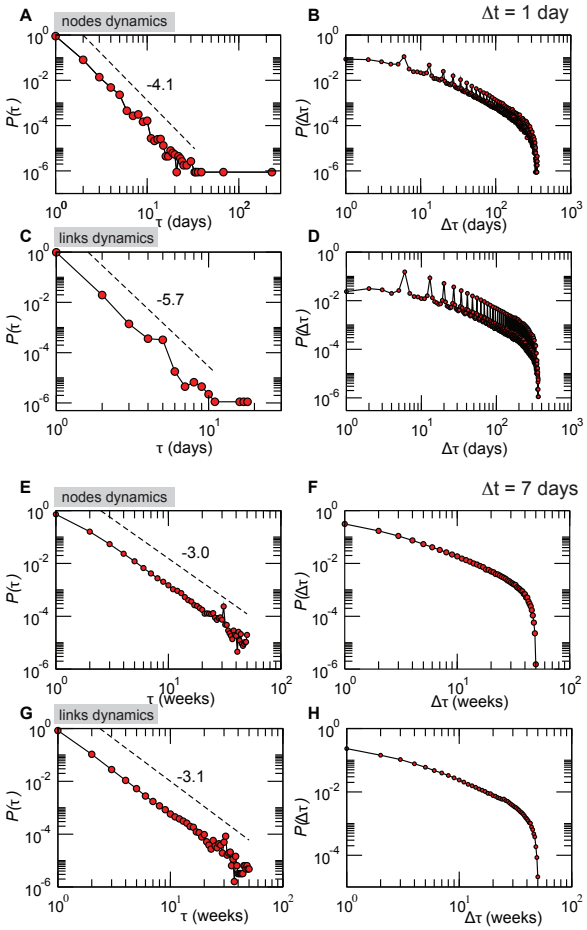

Figure 6.9: Probability distributions of the duration τ of activity and of the duration $\Delta\tau$ of inactivity of nodes and links. Results are reported for daily (panels A to D) and weekly (panels E to H) networks. In the daily case, weekend breaks are neglected as they are characterized by a much lower activity and clear weekly patterns (see Figure 6.3). The observed peaks in $P(\Delta\tau)$ of the daily networks correspond to inactivity periods of multiples of a week.

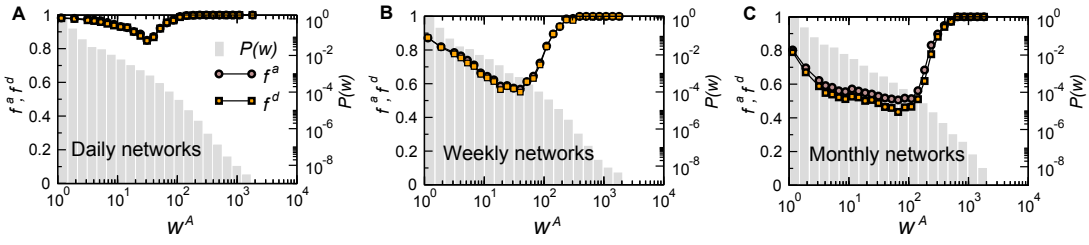

Figure 6.10: Fraction of appearing/disappearing links as a function of the weight associated to the link. The weight considered here counts the number of animals, w^A. Results for daily, weekly, and monthly networks are shown (panels A, B, C, respectively). As a reference, the weight distribution is also shown with a grey histogram.

to the definition of f^d by considering the links of weights w active at time $t-1$ but no longer active at time t. The quantities $f^a(w)$ and $f^d(w)$ are shown in Figure 6.10 for daily, weekly and monthly networks. We observe that $f^a(w)$ and $f^d(w)$ have an almost identical behavior, though dependent on the timescale Δt. Disaggregating the results by premises type for the origin (or for the destination) of each considered link, it turns out that the behavior observed in Figure 6.10 results from a convolution of trends that are quantitatively different but qualitatively similar for all farm types. In all cases, links with small or large displacements of animals are both very unstable, whereas the most stable links are those with an intermediate weight. While till now the system of bovine movements showed properties that are very similar to those found in the analysis of human mobility by air travel, this result instead strongly differs from the positive correlations of links' stability and weight found in the airline transportation network (49). In the airline system this is due to the fact that links with large weights correspond to busy routes that are economically convenient carrying a large fraction of the traffic and thus well established. Different commercial driving forces characterize the cattle trade flows and, in addition, premises have limited receiving capacities, constrained by the limited size of the space hosting the cattle for a widely varying number of days (see the results in Figure 6.2). Since a large weight corresponds to a transport of a large number of animals, it is rather unlikely that two (or more) very large such events occur on the same connection in rapid succession, as this may correspond to a large increase in the population at the premises, if no animals are moved away. Differently from this process in which bovines stay at the arrival node after displacement, airline passengers either connect through an airport or leave the airport to reach their final destination, without thus increasing the population at the mobility node itself. The result is that large displacements are

very stable in the airline case, whereas they heavily fluctuate in the bovine case. The lack of possible identification of stable connections over time carrying large weights (and thus having a large spreading potential) seems to indicate the absence of a robust pattern of movements in the system that could be easily targeted by intervention measures aimed at controlling and containing the spread of a disease. This aspect will be explored in further detail in the next subsection when evaluating the evolution dynamics of the network backbone. As expected, the minimum values of $f^a(w)$ and $f^d(w)$ are very close to 1 when considering the daily networks, meaning that more than 80% of the links present at a given day will disappear the day after (and similarly for the appearance of links). At such timescale the full dynamical nature of the network emerges. More stable structures are instead detected at larger aggregation times, when weekly and monthly networks are considered.

6.4.2 Fluctuations of nodes and links properties

In addition to characterizing the dynamics with which nodes and links can switch on and off their activity, here we study the evolution of nodes' and links' properties while they are active. In particular, the evolution in time of a link's weight $w_{ij}(t)$ is characterized by its growth rate $r_{ij}(t) = log(\frac{w_{ij}(t+1)}{w_{ij}(t)})$ whose distribution is shown (for the weights w^A) in Figure 6.11 for the various time windows under study. The distributions are stationary, with exponentially decaying tails, as found for the airports network (49) and in studies of firm growth (283). This corresponds to a weights' evolution from one month to the next of the form $w_{ij}(t+1) = w_{ij}(t)(1+\eta_{ij})$ where the multiplicative noise $\eta = e^r - 1$ is a random variable whose distribution is broad and does not depend on time, indicating that most of the weights increments are small but that sudden and large variations of the weights can be observed with a small but non negligible probability. The highly dynamical nature of the network, characterized by large instabilities and timescales describing the appearance and disappearance of nodes and links, is expected to have a strong impact on the nodes' properties as well. For instance, a node with many connections on a certain day may be much less connected the next day (284). The stationarity of the distributions obtained from the analysis of static aggregated networks does not imply the stationarity of the properties of each given node; the set of nodes in the tail of the distribution may for instance differ from one snapshot to another. If we focus on properties of centrality of the nodes, which are often used to identify and target the elements of the system for isolation and quarantine aiming at prevention and control of an epidemic spreading on the network, large fluctuations in these values point to the strong limitations of such measures. In order to investigate this, we show the variations of a

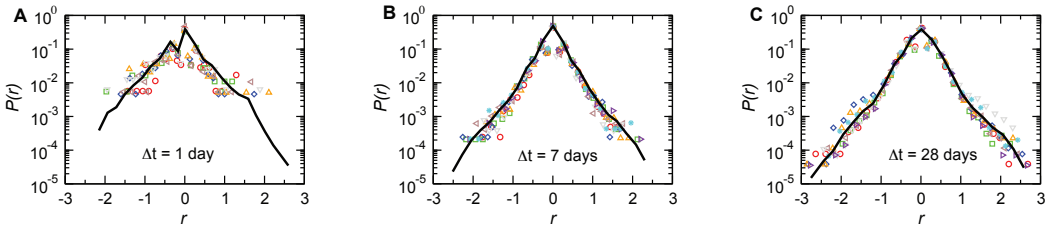

Figure 6.11: . Distributions of the growth rates of the number of bovines w^A displaced along a connection. The solid line represents the distribution of the growth rates considering all networks of a given aggregating time window Δt. Symbols correspond to a selection of snapshots.

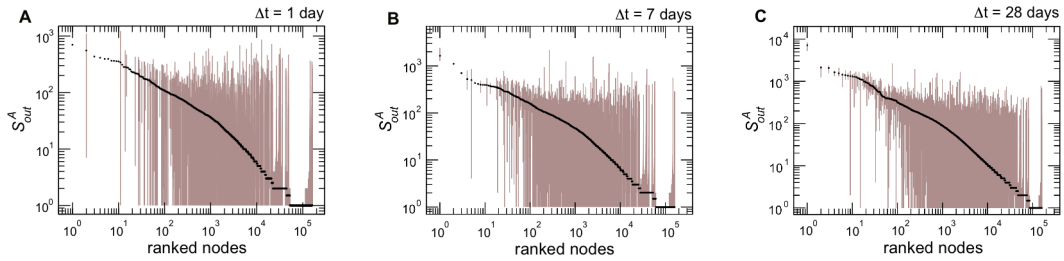

Figure 6.12: Fluctuations of the total outgoing traffic of bovines of a given holding for various aggregating time windows. Each plot shows, for each holding of the system, the fluctuations of the values of s^A_{out} assumed by each node during all snapshots of the Δt under study. The median (black dots) and the 95% confidence interval (brown shaded area) of outgoing traffic are shown.

node's property for all snapshots considered, depending on the timescale Δt under study. Figure 6.12 shows the median and the 95% confidence interval of all values of the out-strength s^A_{out} that each node assumes when active, for different time window lengths. Very large fluctuations are observed, with most nodes showing variations over more than 2 decades, signaling that this property lacks stationarity at the node level. Some nodes with very high strength seem to have no fluctuations, but they appear in fact only once in the dataset. Similar results are obtained when considering other possible measures of node centrality, such as the in-strength or the in- and out-degree (not shown). Given that these quantities are proxy measures for the centrality of nodes, such findings strongly undermine the efficacy of traditional measures for epidemic control that do not take into account the large variations in time of the role of the premises with respect to the flows of the system.

6.4.3 Evolution of the network backbone

The results of the previous subsection show how the system is characterized by large fluctuations and strong topology and traffic variations on all spatial and temporal scales. The overall picture is thus one of a network whose structure changes very strongly from one snapshot to the next, not only at the global level, but also at the node neighborhood and node levels, inducing very strong centrality fluctuations. Notably, these centrality fluctuations are observed for all premises and geographical positions. A natural question therefore arises concerning the possible existence of a backbone of nodes and connections, carrying the relevant topological and dynamical information of the system, and of its temporal stability. The observed strong fluctuations may indeed be related to less meaningful connections of the network, and may thus be compatible with a stationary backbone. A first attempt at defining a global backbone over time consists in considering the intersection of successive aggregated networks. If a considerable fraction of the system is stable across time, the intersection will be quite large and will identify the subset of premises and flows that have a predominant role in the dynamics. At the monthly scale, less than 4% of the links are common to the 13 corresponding networks. Moreover, the corresponding weights are not particularly stable and show a growth rate distribution comparable to the one obtained for the original networks. More advanced filtering techniques can be used to extract a statistically significant subgraph that carries a significant part of the traffic. In particular, it is possible to retain only the links with weights larger than a certain threshold, i.e., the ones with most traffic. However, when the system is characterized by large fluctuations of weights, and more in detail by a large heterogeneity of weights around a given node (as is the case here, not shown), global thresholding can lead to misleading dismissal of locally very important links (285). For this reason, in addition to the thresholding method, we consider the disparity filter method that was introduced in Ref. (285). For each node, it consists in identifying the links that should be preserved in the network. To this aim, one considers the null hypothesis of a random assignment of the normalized weights $p_{ij}^{out} = w_{ij}/s_{i,out}$ and $p_{ij}^{in} = w_{ij}/s_{j,in}$ (as links are directed, we consider two normalized weights definitions), and computes for each link the probability $\alpha_{ij}^{in(out)}$ that its normalized weights are compatible with this hypothesis. These probabilities are given by $\alpha_{ij}^{in} = (1-p_{ij}^{in})^{k_{j,in}-1}$ and $\alpha_{ij}^{out} = (1-p_{ij}^{out})^{k_{i,out}-1}$ (286), and the backbone is given by the links which satisfy at least one of the conditions $\alpha_{ij}^{in} < \alpha$ or $\alpha_{ij}^{out} < \alpha$, where α is a parameter that can be tuned in order to change the significance level of the filtering. For networks with uncorrelated weights, the disparity filtering procedure is equivalent to the

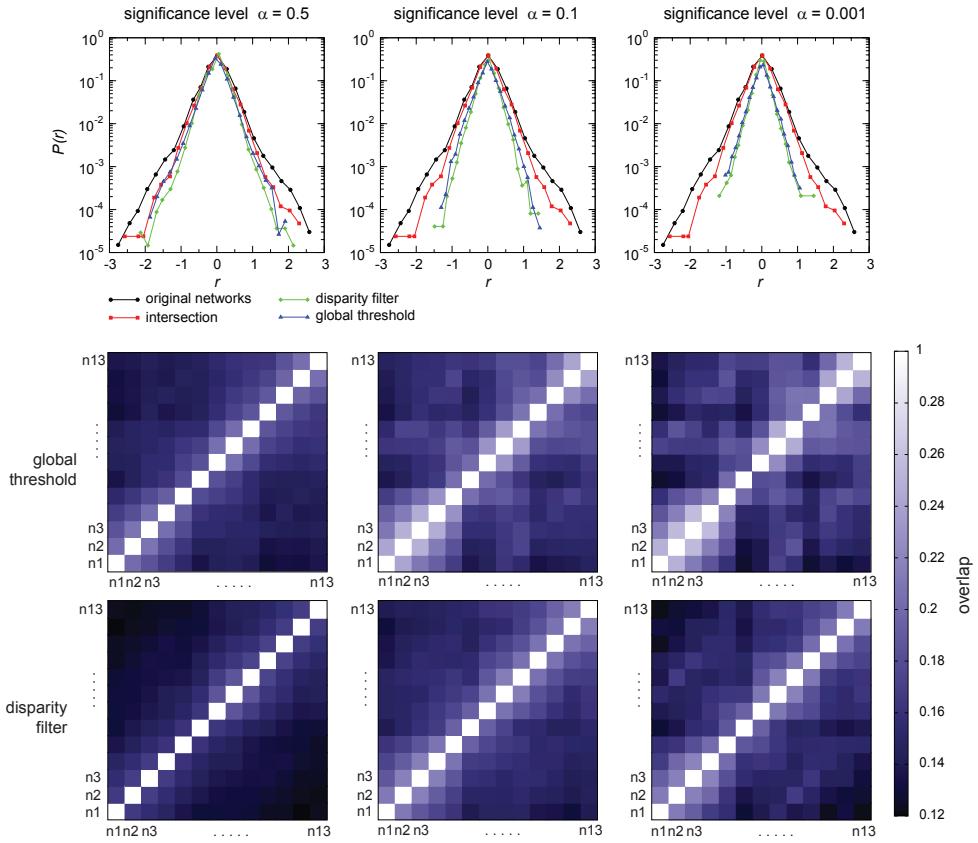

Figure 6.13: Evolution of monthly network backbones. Top: Distributions of the growth rates of the weights w^A of the backbone links, where the network backbone is obtained under different filtering procedures. In each case, growth rates r are measured only for links that are present in two successive backbones. Center and Bottom: Overlap between the backbones of monthly networks. The overlap measures the number of links common to the pair of networks under consideration, normalized by their total number of links. Backbones are obtained either with a global threshold filter (center row) or using a disparity filter (bottom row). Three values of the significance parameter α are considered.

global thresholding procedure, pruning the links with a weight smaller than $\langle w \rangle \cdot ln(1/\alpha)$, where $\langle w \rangle$ is the average weight in the network. We have considered both filtering procedures for the aggregated networks with $\Delta t = 28$ days, and constructed for each network the corresponding backbones for various significance levels. In order to assess how the network backbone change in time, we have then computed the overlap between the $13 \cdot 12/2$ pairs of backbones, at a given significance level, where the overlap of two networks with respective sets of edges E_1 and E_2 is defined by $|E_1 \bigcap E_2|/|E_1 \bigcup E_2|$. Figure 6.13 displays the distributions of the growth rates of the links' weights w^A in the various backbones, compared with the corresponding distribution in the whole network, together with the overlaps of the monthly backbones in color-coded matrices. The overlap between backbones of successive monthly networks is substantial but not large, approximately ranging from 25 to 30%. If we assume that the two successive backbones have the same size, an overlap of $25-30\%$ would correspond approximately to an intersection of $40-50\%$ between the two systems. While about half of the system is retained from one snapshot to the following, this value becomes rapidly smaller when moving away from the diagonal, i.e., as the corresponding networks are further apart in time. This shows that the memory of the most significant links in a given month rapidly fades away in the successive months, and that evaluating the importance of a link based on previous evidence could thus be misleading.

6.4.4 Dynamical motifs

A static representation of the system and the usual tools devised to investigate static networks are clearly insufficient to properly describe and analyze the temporal dimension of the dataset. After observing the large fluctuations and the fast dynamics characterizing the system at all timescales, we present in this subsection an analysis aimed at going even further in the understanding of the system flows by exploring the possible signatures of a temporal ordering of the bovine displacements and the presence of recurrent paths. The search in networks of the abundance of particular topological paths or motifs (287) should then be complemented by causality requirements and approaches that are able to incorporate the longitudinal dimension (70; 288; 289; 290; 291). Beyond the topological paths, the temporal correlation is fundamental for the study of diffusion processes. One of the main consequences of the temporal evolution of the network resides in the causality constraints it induces. For instance, a spreading phenomenon can propagate on a path ijk (i.e. from i to j to k) only if the link ij is present before the link jk (70; 292). This becomes particularly relevant if the flows form cycles or paths that allow the re-infection of some premises, given an appropriate interplay of the disease and movement timescales. From the point of view

of quarantine and similar control strategies, this would represent an important phenomenon to take into account when establishing the identification of premises to isolate, or the durations of disease surveillance at those locations. Figure 6.14 presents an example of causal motifs: for instance, the repetition of the sequence of a link ij followed in the next snapshot by a link jk could imply a cause-effect relationship between these two links. Here we introduce a new measure to define causal motifs and restrict our analysis on the shortest possible timescale, i.e. the intrinsic timescale of the system $\Delta t = 1$ day. We collect, for each path length , the motifs given by a list of links $i_0i_1, i_1i_2, , i_{l-1}i_l$ such that i_0i_1 is present at a certain snapshot t_0, i_1i_2 at snapshot t_0+1, and finally $i_{l-1}i_l$ at $t_0 + l - 1$. The duration of the path is therefore equal to its length, and each path corresponds to a possible propagation that respects causal constraints. Each motif can occur for several values of the starting time t_0, and motifs of a given length can be ranked according to their number of occurrences. It is worth remarking that the above definition does not focus on the shape of the motifs since only temporally connected chain-like motifs are considered, and the recurrence is sought at the microscopic level counting the number of appearance of a certain link sequence. Figure 6.14B shows the corresponding frequency-rank plots, as well as the fraction of motifs that are repeated more than once for each length. Motifs are found up to length $l = 8$, and both the absolute number of motifs and the fraction of recurring motifs strongly decrease for increasing lengths. Since the number of times a link is present in the daily networks is broadly distributed (this number is given by the weight w^B in the globally aggregated network over the whole year), pure statistical effects could be responsible for the abundance of specific patterns. For instance, if both links ij and jk are present all the time, then the causal path ijk will be very frequent. We therefore compare in Figure 6.14C the results obtained in the real data with different null models. The first null model (random ordered) is constructed by randomly shuffling the order of the daily aggregated networks: in this way, the structure of each daily network is kept, but the temporal correlations are lost. The second null model (temporal mixed edges) shuffles randomly the days in which each edge is active, independently from one edge to the next. The resulting daily aggregated networks have therefore randomized structures. Finally, we construct also a third null model by reshuffling the edges in each daily network as described in Ref. (293) (time ordered and reshuffled networks): we recall that this procedure consists in taking at random pairs of links ij and lm involving 4 distinct nodes, and rewiring them as e.g. im and jl (if neither im nor jl already exist) . This procedure preserves both the in- and out-degree for each node, but destroys correlations. Figure 6.14 shows that the two first null models lead to similar results: a much smaller number of motifs is observed, and

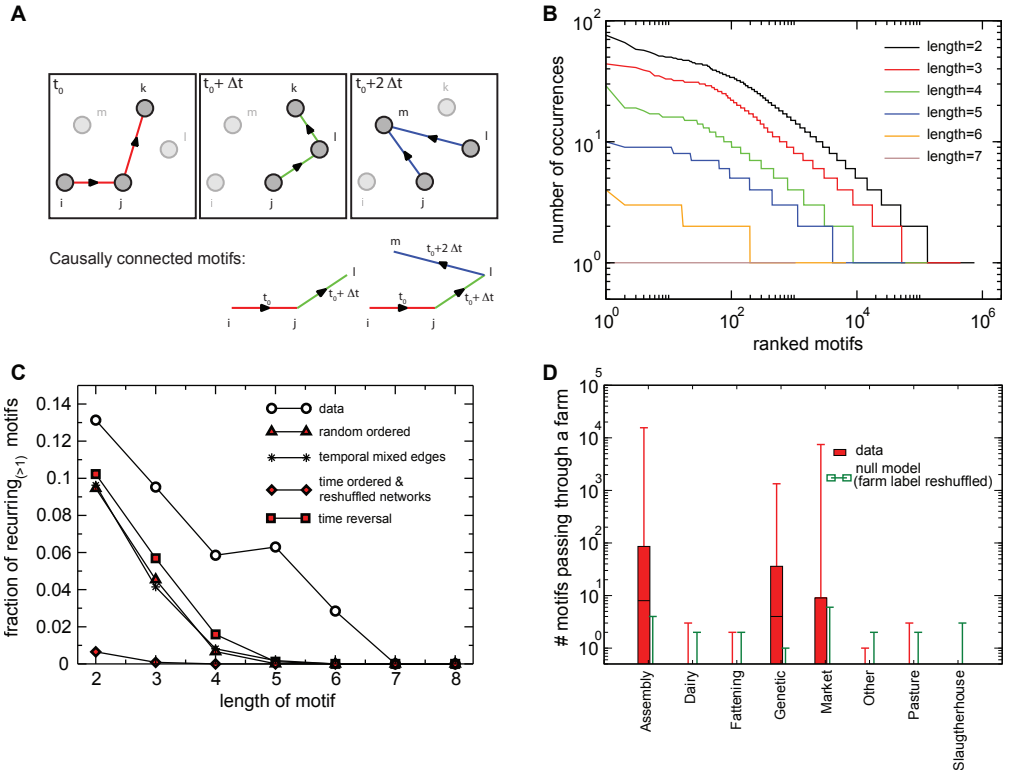

Figure 6.14: Motifs: schematic representation and their occurrence. A schematic example of the dynamics of a subset of the mobility networks is shown in panel A through three successive snapshots. The connections are color-coded according to the time at which they are active. A temporal motif is a temporal sequence of links such that the destination node of a link at time t_0 is the origin of another link at time $t_0 + \Delta t$. Two examples of motifs, of respective lengths 2 and 3, are shown below. We restrict the present study to the case of $\Delta t = 1$ day. Panel B shows the results on the presence of motifs, analyzed by counting the number of occurrences during the timeframe under study. The longer the motifs, the smaller the number of times they appear. By focusing only on the set of motifs that occur at least twice, panel C compares the size of this set (expressed as a fraction of the total) obtained from the empirical dataset with the sizes obtained through various randomization procedures. The results are shown as functions of the motifs length. In panel D the median and confidence intervals of the number of motifs passing through a farm depending on the farms type are shown, together with the same computation for a null model in which the farm types are reshuffled at random.

a smaller fraction of these motifs are found more than once. At lengths smaller than 5 however, this fraction is non negligible, showing that purely statistical effects due to the frequent presence of some links account for a part of the motifs presence and repetition. When both time ordering and network topology are reshuffled, motifs essentially disappear.

Since the network under study is directed, it is interesting to note that a causal sequence of links ($i_n i_{n+1}$ at a certain snapshot t_n, followed by $i_{n+1} i_{n+2}$ at snapshot $t_n + 1$) is not a valid causal path if it happens in the reverse order ($i_{n+1} i_{n+2}$ followed by $i_n i_{n+1}$). We therefore consider in Figure 6.14 also the sequence of 365 daily aggregated networks, seen in the reverse temporal order. Strikingly, the number of motifs is much smaller than for the true temporal sequence, and the fraction of repeated motifs is close to the case of a random temporal ordering. This indicates the presence of an intrinsic time arrow in the dataset, and provides a general method for investigating this aspect in dynamically directed networks. To our knowledge, this is indeed the first time that an intrinsic arrow of time has been explicitly detected in a temporal network. In Figure 6.14 we also show the statistics of the number of motifs passing through a farm for different farm types. In order to take into account the relative abundance of the different farm types, we compare the results with a null model where the labels describing the farm types are reshuffled. We notice that some premises types (such as assembly centers or markets) are much more prone to be part of causal motifs than what would be expected for a random labeling of the premises. Our definition of causal motifs is therefore able to characterize the behavior of premises by identifying those types of premises that, as expected, show highly recurrent flow-in/flow-out patterns at such short timescale. The present analysis can also be extended by considering longer latency times for the occurrence of specific causal paths in the network, by considering sequences of links ij at time t and jl at time $t + t'$, relaxing the previous condition on the separation of times between the occurrences of successive links in the motifs. The flexibility of this approach thus allows the tuning of the analysis to the relevant timescales of the dynamical process under study, with a variable latency time t' that corresponds to the time during which a node can be considered as continuously active.

6.5 Spreading processes on dynamical networks

Surveillance, prevention and control represent major aspects of the public health response to a potential contagious disease (294). They aim at monitoring the health status of the population of hosts, detecting and observing the progression of an outbreak, as well as devising and

implementing effective measures to best minimize its impact on the population. While the spatial and topological aspects of the pattern of contacts among hosts have been widely explored in the epidemiological context, we now focus on a dynamical framework in which the temporal nature of the contacts is considered. In particular, the dynamical behavior of the network substrate observed in many real-world examples (ranging from social interactions to mobility patterns (21; 266; 288; 295; 296; 297)) opens theoretical and computational challenges concerning: (i) the lack of meaningful definitions of nodes' vulnerability, given the potential large fluctuations of measures of centrality defined on static structural properties, and (ii) the strong dependance of the spreading pattern on the initial conditions. Here we address these challenges first through a percolation analysis of the cattle displacement network and then by performing disease spreading simulations among livestock premises where the full temporal resolution of the system is considered.

6.5.1 Percolation analysis

The strong temporal variations in the network network topology higlighted in Figure 6.12, along with the very short memory of the backbone structure (see Figure 6.13) leads us to the study of how the dynamical aspects impacts the percolation properties of the network of displacements. Percolation has long been used in the analysis of complex networks (298; 299), and results have shown that many real-world network structures typically retain their integrity, in terms of global connectedness, when nodes or links are removed in a random fashion, while they are very fragile with respect to targeted attacks. In this respect, percolation analysis has become a tool to investigate the structure of networks, by studying how the size of the largest connected component evolves when nodes are removed according to different procedures (298; 299; 300; 301; 302). The size of the giant component not only is a measure of the resilience of the structural properties of the network under study, but it also quantifies the extent to which an epidemic could possibly spread in the system. Identifying ways to reduce this size, by removing particular nodes, is equivalent to finding efficient intervention and control strategies in the framework of disease spreading, aiming at breaking down the network in small pieces in order to prevent the disease from invading the system. Let us consider for instance that an outbreak starts at a certain date. It is then possible to sort the nodes of the aggregated network of the corresponding time window by their degree, strength, or other centrality measures, and to try to contain the disease spread by isolating the most central nodes, reducing drastically the size of the largest connected component through the isolation of only a few percents of the nodes (298; 300; 301). A lot of

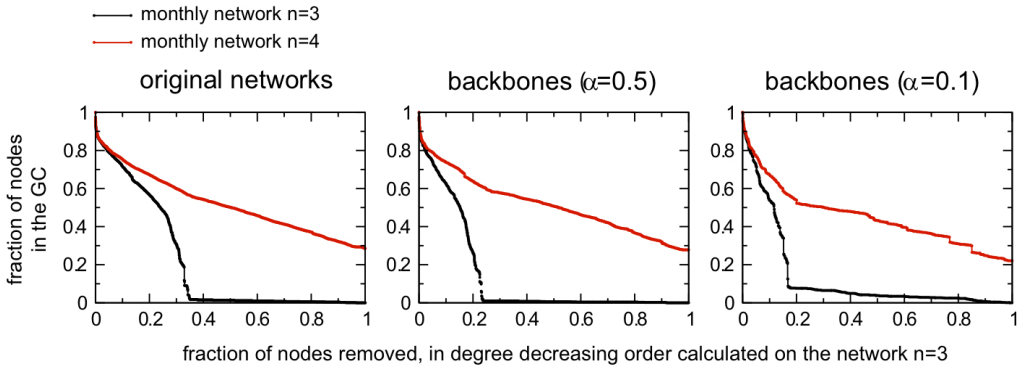

Figure 6.15: Percolation analysis on consecutive monthly networks. Two consecutive monthly snapshots ($n = 3$ and $n = 4$) have been considered. A list of nodes with decreasing degree is calculated on the snapshot $n = 3$, and is applied as a removal strategy for both networks. The same procedure has been performed on the corresponding network backbones obtained for two values of the significance parameter α.

work has been done in this direction for the analysis of the fragmentation of the network of livestock movements when nodes are chosen according to different centrality measures (256; 258; 259; 260; 261; 265; 267), and no information on disease spreading is considered, as instead was done in Ref. (303). However, these studies have neglected the dynamical nature of the system, focusing on specific snapshots only, and assuming to be able to access all the relevant information of the system at any given point in time, e.g. during an epidemic emergency. Since we showed so far how the underlying topology and flows strongly fluctuate at all levels, here we want to study instead the situation in which we have limited information on the system gathered from its activity on the last time window under study, and we want to apply isolation and quarantine measures to the following snapshot. The ranking of nodes according to a given centrality measure (corresponding to their spreading potential) computed on a certain time window may indeed loose its relevance when applied at successive times. Given these intrinsic dynamical features, we aim here at assessing the impact of a removal strategy on consecutive snapshots (thus the snapshots characterized by the highest values of the overlap), once the strategy is defined on the basis of the available information on one snapshot only, and is not updated according to the successive network evolution. We investigate this aspect by measuring the effect of the successive removal of nodes by decreasing degree in consecutive snapshots of $\Delta t = 28$ days. More in detail, by focusing on a given snapshot for $\Delta t = 28$ days (the third snapshot of the

year, chosen as an illustrative example), we fix the order of nodes to be removed in a degree-decreasing fashion. Then, we assess the impact of the removal of nodes ordered in such way on this snapshot and on the following one. This means that for the successive snapshot we are not re-evaluating the centrality of each node (as measured here by the degree) but we use the information computed on the previous time snapshot. This procedure is tested on the full network and on the corresponding backbone, calculated at two different significance levels. Figure 6.15 shows the results in terms of the relative size of the giant component as a function of the fraction of nodes removed. As expected, the removal of nodes is very efficient if the order of nodes to be removed is calculated on that snapshot (256; 258; 259; 260; 261; 265; 267), whereas such ordering is not able to destroy the network at the successive time window, leading to a size of the giant component that decreases very slowly, and maintains a fraction of more than 20% of the system still intact and connected after the list of nodes is exhausted. Even though a large number of nodes is removed from the system, the effectiveness of such isolation procedure is strongly limited by that fact that the premises' properties have dramatically changed. Many of the active premises have appeared/disappeared from one snapshot to the other, and the ones that remained have strongly changed their interaction pattern. In such situations, intervention and control strategies devised using the information from static aggregated networks, or more generally from data from past mobility patterns, can thus result to be very inefficient.

6.5.2 Epidemic spreading simulations

In order to gain a general understanding of the interplay between the dynamical evolution of the disease and the temporal features of the animal movements, we consider a simple model of a notifiable disease characterized by short timescales. The disease spread on the dynamical network is modeled using a simple SIR compartmental model (71). We assume that farms are the discrete single units of the process, following a Levins-type approximation for metapopulation systems (304) that ignores the possible impacts of within-farm dynamics, as commonly assumed in the study of the spread of zoonosis through animal movements (305). Premises are labeled as Susceptible, Infectious, or Removed, according to the stage of the disease, and are not further distinguished according to their characterization or function (e.g. markets, fattening farms, or others), as we aim at exploring the role of initial conditions in defining a propagation pattern on the dynamical network of movements, independently of the farm type. All farms are considered susceptible at the beginning of the simulations, except for the seeding farm. At each time step, an infectious farm i can transmit the disease along its outgoing link to its neighboring susceptible

farms that become infected and can then propagate the disease further in the network. Here we consider a deterministic process for which the contagion occurs with probability equal to 1 as long as there is a directed link of cattle movements from an infectious farm to a susceptible one at a given time step. Though a crude assumption, this allows us to simplify the computational exploration of the initial conditions, define and evaluate the proposed methodology for the "simple" deterministic case, focusing on the fastest infection patterns and yielding an upper bound to the final size of the epidemics. The corresponding stochastic case is then discussed in the last section. After μ^{-1} time steps an infected farm becomes recovered and cannot be reinfected. The simulation is fully defined by the choice of the timescale Δt, used to define the successive aggregated networks on which the disease propagation occurs, and of the initial conditions (x_0, t_0) where x_0 is the seeding node (i.e. the first infected farm) and t_0 is the seeding time (i.e. the time at which x_0 starts being infectious). The choice of the aggregating time window length Δt affects the underlying mobility structure, leading to denser displacement networks for longer time windows, while the time step used in the numerical simulations of the spreading dynamics is kept fixed to 1 day. In this perspective, when the spreading process takes place on the daily dynamical networks, at every time step of the spreading the snapshot of the static network is different, while for a longer aggregating window length the network topology remains unchanged for exactly Δt time steps. In Figure 6.16 , the unfolding of the spreading for different aggregating time window lengths is followed by plotting the temporal evolution of the number of infected premises for every spreading time step (=1 day). We first explore the role of the timescale of aggregation Δt of the dynamical network on the disease propagation, by analyzing the spreading patterns resulting from outbreaks starting at each of the $\sim 1.7 \cdot 10^5$ premises on the seeding date t_0 = January 1st, assuming an infectious period $\mu^{-1} = 7$ days.

The simulated epidemics dramatically depend on the aggregation timescale, as shown by the time behavior of the number of infected farms for outbreaks taking place on daily, weekly, monthly and yearly aggregations of the displacements, and for all the possible origins x_0, as reported in Figure 6.16. As Δt increases, the spreading becomes faster and reaches a larger proportion of the nodes (266). At short timescales of aggregation, the large temporal variability of the displacement networks limits the spreading of the disease, by creating disconnected components that prevent possible paths of infection along the network. With increasing values of Δt, this intrinsic variability is collapsed into an aggregated static picture, where links active in reality at very different times are simultaneously present in the same static network, thus allowing propagation paths that were otherwise prevented for shorter Δt values. The choice of the aggregating

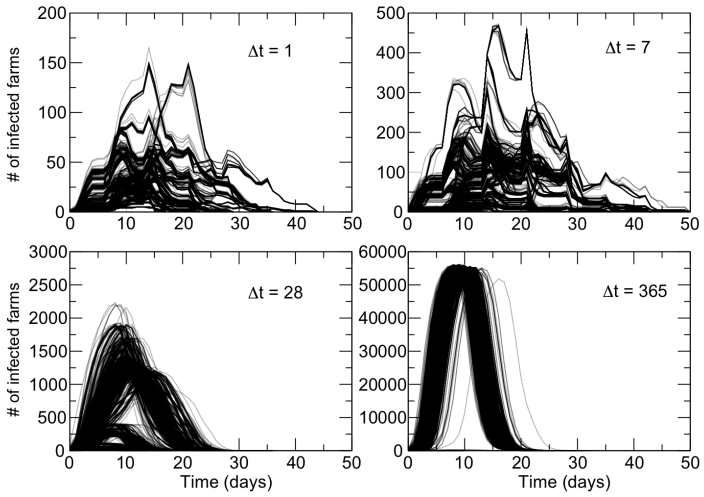

Figure 6.16: Number of infected farms as a function of time for different aggregating time windows Δt. Each grey curve represents the profile of an epidemic starting on January 1st from a given seed.

time window length Δt affects the underlying mobility structure, leading to denser displacement networks for longer time windows, while the time step used in the numerical simulations of the spreading dynamics is kept fixed to 1 day. In this perspective, when the spreading process takes place on the daily dynamical networks, at every time step of the spreading the snapshot of the static network is different, while for a longer aggregating window length the network topology remains unchanged for exactly Δt time steps. Besides overall differences in timing and size reached by the epidemic with changing Δt, the epidemic profiles also show an intrinsic variability that depends on the initial conditions of the outbreak, and that is averaged out by large values of the aggregation timescale. When the maximal temporal resolution is used (i.e. $\Delta t = 1$ day), the profiles of Figure 6.16 are rather ruffled, with multiple peaks and strong differences in peak times for different initial conditions. Larger Δt values lead to smoother curves that collapse into a small set of synchronized behaviors of the global spread originating from different seeds (e.g. for $\Delta t = 28$ days), with the number of different profiles being reduced by the aggregation on longer time windows, as shown by the extreme case of $\Delta t = 365$ days, where no temporal fluctuations are present and changes in the profiles only depend on changes in geography of the initial conditions. By loosing the intrinsic variability at the local level due to the aggregation on large Δt values, also the role of the initial conditions in leading to outbreaks of different nature (in terms of size and timing) is increasingly weakened. Therefore, in order to realistically account

for the impact of the seeding on the spread of fast epidemics on the dynamical network, in the following we focus on the finest temporal grain, $\Delta t = 1$ day, for the description of the bovines mobility in the epidemic simulations.

6.5.3 Invasion paths and seeds' cluster detection

Given the limited applicability of quantities defined *a priori* to characterize the spreading potential of a node in such a highly dynamical network, here we exhaustively explore the dependance of the spreading process from the initial conditions and investigate the possible emergence of recurrent patterns, aimed at proposing a method to identify similar spreaders in such a complex environment. The spreading pattern of an outbreak is encoded in an invasion path where a directed link $i - j$ denotes that the farm j has been infected by the farm i (78). Each invasion path Γ is thus composed by an ensemble of nodes $\vec{\nu}$, an ensemble of links $\vec{l}$, and a root x_0 representing the initial seed of the outbreak. We define the overlap Θ_{12} between two paths Γ_1 and Γ_2 as the Jaccard index $\frac{\vec{\nu}_1 \bigcap \vec{\nu}_2}{\vec{\nu}_1 \bigcup \vec{\nu}_2}$, measuring the common nodes with respect to the total number of nodes reached by the two paths. It is important to note that this measure does not consider the information on the links of transmission from one farm to another, as we are indeed interested on the observable outcome of the outbreaks, namely the fact that a farm is infected or not, rather than on the precise path followed between the infected nodes. We compute, at fixed $\Delta t = 1$ day and initial time t_0, the overlap Θ_{12} between the invasion paths of deterministic SIR outbreaks generated by every pair of potential seeds (x_1, x_2). We then construct the initial conditions similarity network (ICSN) as a weighted, undirected network in which each node is an initial condition of the epidemic spread and the link between two nodes x_1 and x_2 is weighted by the value of the overlap Θ_{12}, measuring the similarity of the invasion trees they produce (306). By filtering the ICSN to disregard too small values of the similarity, below a given threshold Θ_{th}, the network separates into several connected components, leading to a natural classification of the initial conditions into clusters. These represent sets of nodes that, if at the origin of an outbreak, would lead to similar invasion paths. Clusters are organized in a hierarchy depending on the value of Θ_{th}, and it is interesting to note that, given the distribution of similarity values obtained, even large enough values of Θ_{th} lead to the emergence of non-trivial clusters of initial conditions, i.e. different from simply isolated nodes. As the choice of the threshold is arbitrary, it is important to check the robustness of the obtained cluster structure with respect to changes in the threshold value. We investigate this point in Figure 6.18 by measuring the intersection of clusters obtained with different threshold values. For large enough threshold values, the structure

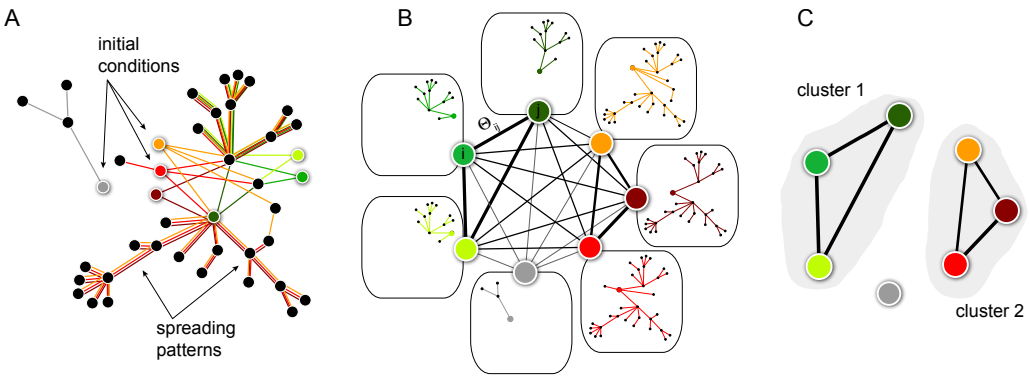

Figure 6.17: Schematic representation of the cluster detection procedure. (A) Different simulated invasion paths (colored lines) obtained for different seeder (colored nodes) are shown on the network. Each path is colored according to with the color of the seeding node. (B) The initial conditions similarity network (ICSN) is obtained by calculating the overlap Θ_{ij} for any pair of initial conditions i and j, measuring the similarity between the invasion paths originated by the two nodes. Thicker lines in the ICSN indicate a higher overlap. (C) By filtering the ICSN neglecting all links characterized by an overlap lower than a given threshold Θ_{th}, two non-trivial connected components emerge (i.e. with a size larger than one farm) corresponding to cluster 1 and cluster 2 of nodes leading to similar propagation pathways.

of resulting clusters is stable with respect to small variations in the thresholding criterion. In all the remaining text we use the threshold value 0.8.

In Figure 6.19, we show the 12 largest clusters identified with this procedure by assuming $\Theta_{th} = 0.8$, along with the displacement network aggregated over the entire spreading period. Some important characteristics of the clusters emerge clearly. First, the nodes of a given cluster defined through the ICSN are not tightly connected in the aggregated displacement network (Figure 6.19A). In addition, there is a lack of chains of infections: the nodes in the clusters are not trivially connected to each other by links that bring the disease from one node to the next. A direct analysis of the aggregated displacement network, based for instance on the search of communities or chain-like motifs, would therefore not be able to detect the similarity of their spreading properties. Most importantly, such topological analysis would not provide meaningful insights for the cluster identification if tested on variations of the timescale of aggregation Δt; the large temporal fluctuations prevent the definition of a unique relevant timescale for the process, and the study of the displacement network aggregated over the entire spreading period (i.e. containing the full information on all displacements occurred) highlight the lack of cluster-like

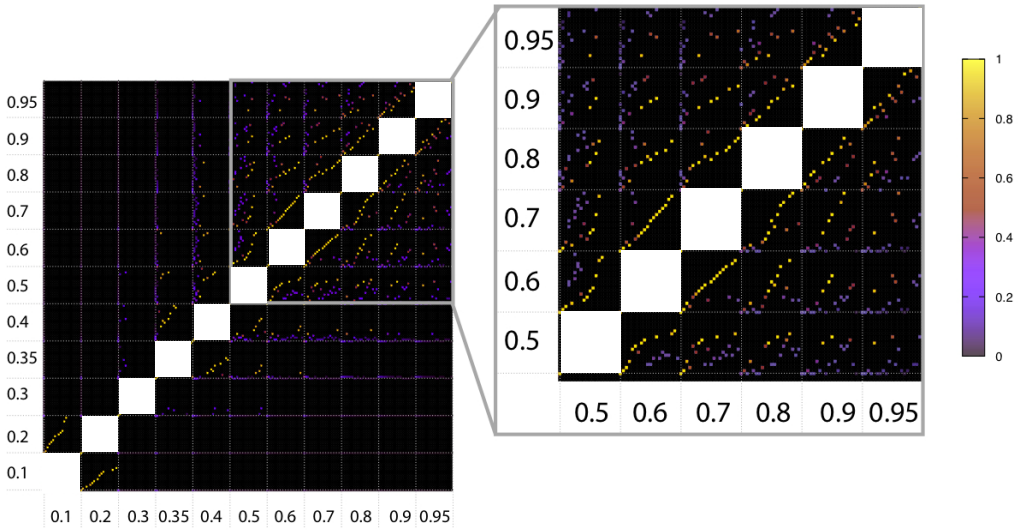

Figure 6.18: Jaccard indices between clusters constructed using different threshold values. For each couple of threshold values a and b, we show at row a and column b a color-coded matrix of the Jaccard indexes between the two sets of 20 largest clusters of the ICSN obtained for the thresholds values a and b. The Jaccard indices of two clusters is computed as the number of common nodes divided by the number of nodes in the union of the clusters. The cases $a = b$ are not shown as they trivially have a diagonal equal to 1 and zero off-diagonal elements. The violet-to-yellow color scale indicate how much the clusters obtained with different thresholds have in common. For threshold values larger than 0.6, the cluster structure is rather stable with respect to small changes in the thresholding criterion. Note that for each threshold value, the 20 largest clusters are ranked by size; as this ranking may change from one cluster value to the next, the yellow dots are not all on the diagonal of the corresponding matrix.

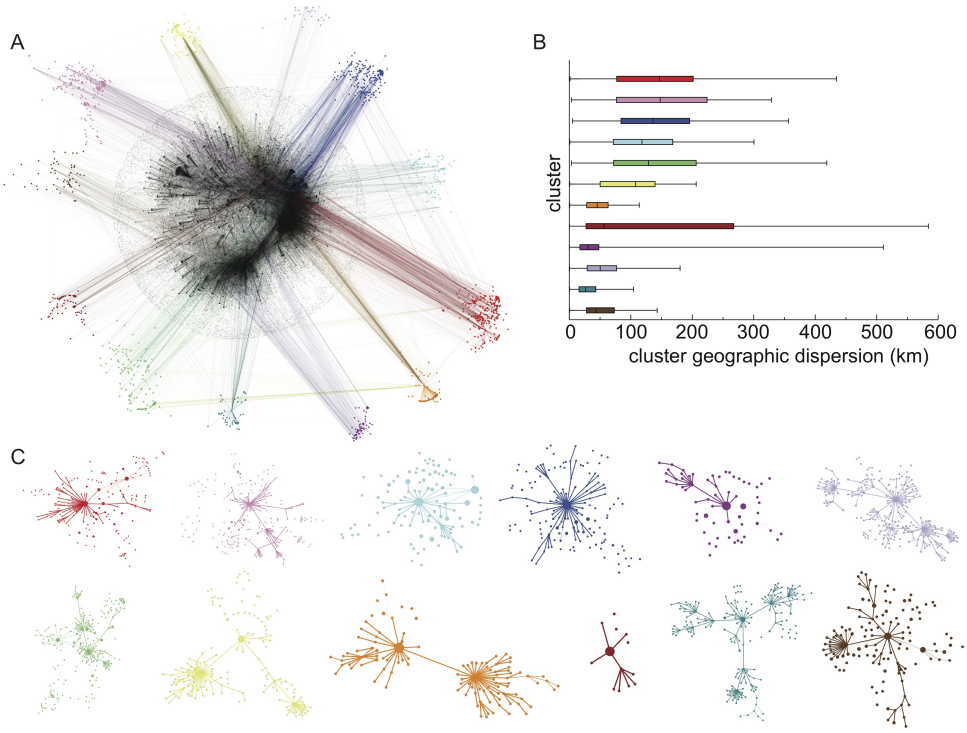

Figure 6.19: Twelve largest clusters obtained from the study of simulated epidemics starting at time t_0 = January 1st. (A) The nodes belonging to each of the 12 clusters are represented in the network of bovines displacement aggregated over the whole spreading period. Each cluster is identified with a color. The nodes in each cluster are not tightly interconnected in this network representation. (B) Cluster geographical dispersion, calculated as the distance between each pair of nodes belonging to the same cluster, and expressed in Kms. The boxplots are colored accordingly to the corresponding clusters. (C) Each network corresponds to a given cluster (indicated by the color) and represents the union of all invasion paths starting from the nodes of that cluster. The initial conditions are not shown in these networks for the sake of simplicity of the visualization; the size of the nodes indicates the number of incoming infections along the invasion paths, and the link thickness represents the number of invasion paths propagating along that connection. Different topological structures of the invasion paths are found for different seed clusters.

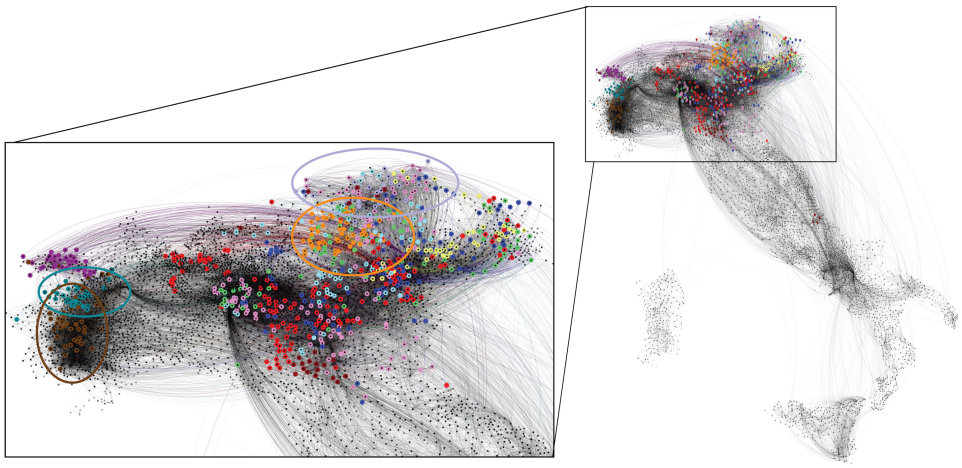

Figure 6.20: The 12 largest clusters are shown in different colors on the georeferenced network of bovines' displacement aggregated over the whole spreading period (35 days). The most compact clusters in terms of geographical dispersion are highlighted with an ellipse.

structures. The geographical analysis of the georeferenced representation of the clusters (where each node is assigned the location of the corresponding municipality) shows that, although some clusters are formed by nodes which are geographically rather close, most clusters are dispersed, with a distribution of distances between nodes spanning several hundreds of kilometers (see panel B of Figure 6.19). Clusters can also geographically overlap and do not have mutually separated geographical boundaries as shown in figure 6.20. Therefore, the geographical proximity of two nodes does not necessarily imply a similar pattern of invasion when the nodes are at the origin of an infectious disease outbreak.

Overall, neither the structural nor the geographical analysis of the dynamical network of displacements would be able to reveal the existence and composition of groups of nodes leading to similar spreading patterns, and a detailed analysis of the dynamical process itself is needed. Interestingly, the mixed shapes observed in the epidemic profiles of Figure 6.16 are automatically classified into a set of specific and well-defined profile behaviors by considering initial conditions belonging to the same cluster, as shown in Figure 6.21A. The clustering method described above, grounded on the comparison of the infected nodes, however disregarding information on the timing of the spreading events, is able to group the spreading histories into similar patterns characterized by the same timing for the overall epidemic and for its peak.

Similar findings are obtained also considering a stochastic infection dynamics, where we gen-

eralize the procedure to a stochastic framework and we recover the results regarding the classification of profile behaviors induced by the clustering, analogously to Figure 6.21A.

6.5.4 Longitudinal stability of the seeds' clusters

The method described above unveils a very interesting partition of the possible seeds of a disease spreading which would not be directly detected from the network structure. The partition however depends a priori on the starting time of the spreading. Given the strong variability of the network's properties on all timescales (297), partitions obtained for spreading processes starting at different times could substantially differ.

In order to investigate this aspect, we define the partition $\mathcal{P}(t_0)$ of the C largest clusters corresponding to the initial time t_0, i.e. $C_i(t_0)$ $(i = 1, \cdots, C)$, and compare it with the partition $\mathcal{P}(t)$ corresponding to a spreading initial time $t > t_0$. In particular, we are interested in measuring to what extent the clusters $C_i(t_0)$ of $\mathcal{P}(t_0)$ are preserved in the partition $\mathcal{P}(t)$. For each cluster $C_i(t_0)$, we thus define the vector $\vec{\rho}_i(t, t_0)$ with components $\rho_{i,j}(t, t_0) = \frac{|C_j(t) \bigcap C_i(t_0)|}{|C_i(t_0)|}$ representing the fraction of nodes of $C_i(t_0)$ present in the cluster $C_j(t)$, i.e. in the cluster j of the partition $\mathcal{P}(t)$. If the cluster detection procedure yields the same clusters at time t_0 and t, each vector $\vec{\rho}_i(t, t_0)$ will have one component equal to 1, and all the others equal to 0. If instead the nodes of $C_i(t_0)$ are homogeneously redistributed into the clusters $C_j(t)$, $\vec{\rho}_i(t, t_0)$ will have all components equal to $1/C$. We restrict our analysis only to the seeding nodes that are possible seeders (i.e., present in the daily network) both at time t_0 and t. Furthermore, since we are considering only the C largest clusters (we will consider $C = 20$ in the following), we note that the fraction $\sigma_i(t, t_0) = \sum_j \rho_{ij}(t, t_0)$ of the nodes of $C_i(t_0)$ also represented in one of the clusters of the partition $\mathcal{P}(t)$ may be smaller than 1. The level of heterogeneity of the time evolution of each cluster $C_i(t_0)$ can be described through the disorder of the vector ρ_i, measured by the normalized entropy function $H(t)$:

$$H_i(t) = \frac{1}{\sigma_i(t) \log(\sigma_i(t)/C)} \sum_j \rho_{i,j}(t) \log \rho_{i,j}(t)\,, \tag{6.1}$$

where we have dropped the dependence on t_0 for simplicity. If $C_i(t_0)$ is also a cluster of $\mathcal{P}(t)$, $H_i(t) = 0$. If its nodes are equally divided into the C clusters of $\mathcal{P}(t)$, the entropy is equal to 1. In general the entropy takes values in the interval $\left[\frac{1}{1-\frac{\log C}{\log \sigma_i}}, 1\right]$, where its minimum value, $\min_{H_i(t)}$, represents the best configuration; all the nodes of $C_i(t_0)$ are in the same cluster of $\mathcal{P}(t)$, except the fraction $(1 - \sigma_i)$ that do not belong anymore to the largest C clusters. Taking into account this quantity is crucial, as high entropy values may occur because the nodes of a cluster $C_i(t_0)$

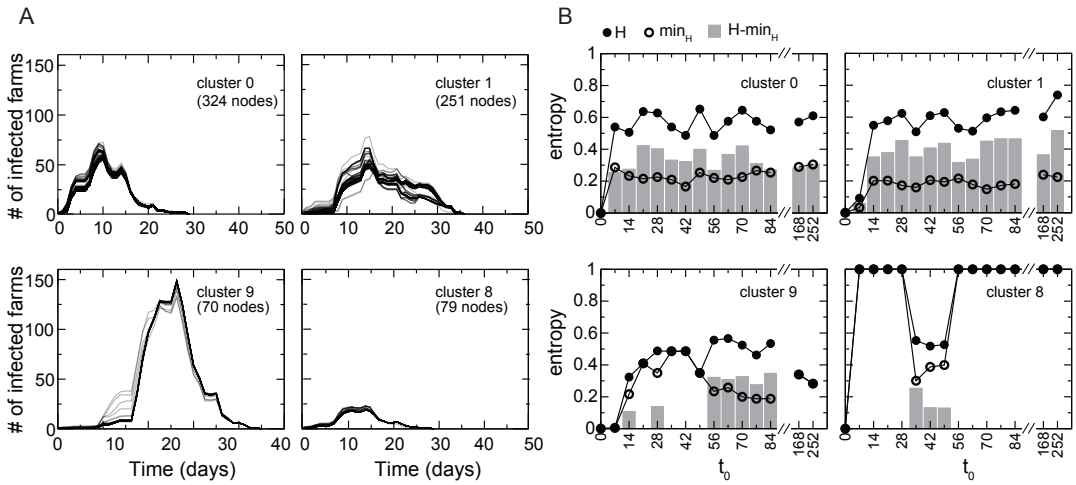

Figure 6.21: Seeds' clusters characterization in terms of epidemic profiles and stability over time. (A) Number of infected farms as a function of time where the curves in each plot correspond to epidemics starting from the initial conditions belonging to the same cluster. Four clusters of different sizes are shown as examples. (B) Entropy of the partition into cluster as a function of time, for the clusters proposed in the panels A. It measures the fragmentation of the largest $C = 20$ clusters obtained for the starting time t_0 = January 1st in the partitions obtained at the following weeks. The difference $H - \min_H$ (grey bars) represents the robustness of the cluster (the smaller the difference and the more robust is the cluster), given that only part of it may be present in the partition obtained for a later starting condition (as measured by $\min_H$). Four typical behaviors can be characterized, each reported through a cluster example: a cluster loosing part of its structure already after 1 week (cluster 0) or after 2 weeks (cluster 2), then showing a stable behavior in time with its nodes partially regrouped in other clusters; a very stable cluster preservation at almost all times (cluster 9); a cluster that is not preserved in time (cluster 8).

obtained at t_0 are no longer active at time t (large $\min_{H_i(t)}$) so that the cluster simply disappears, or because the structure of the network has changed so that, even if the nodes of $C_i(t_0)$ are still potential seeds at t, spreading processes starting from these nodes lead to distinct invasion paths. Measuring the minimum entropy value together with the entropy allows us to distinguish between these two behaviors. As an example, we present in Figure 6.21B the results corresponding to t_0 = January 1st and $t = t_0 + 7w$ with $w = 1, 2, 3, ...$, i.e. successive times separated by w weeks from t_0 (these days being all Mondays to preserve the weekly fluctuations of the network). Our results show that the cluster temporal stability exhibits four main behaviors, pictured in the four panels of Fig. 6.21B: *i*) a substantial fraction of the nodes of the cluster disappears already for $w = 1$ ($\min_H \neq 0$), and small groups of nodes are redistributed in other clusters (small differences $H - \min_H$), with a behavior stable in time (cluster 0); *ii*) a similar behavior,

starting from $w = 2$, since after 1 week the cluster is almost unchanged, showing a very low entropy (cluster 1); *iii*) a very stable behavior at $w = 1$, followed by a robust preservation of the partition (small differences $H - \min_H$) given that a fraction of the nodes of the partition at time t_0 have disappeared from the largest C clusters (cluster 9); *iv*) the cluster is very unstable and the cluster's nodes disappear almost completely from the clusters of the partition $\mathcal{P}(t)$, (very high $\min_H$, cluster 8). The most robust behavior in time, i.e. behavior *(iii)* shown by the example of cluster 9 in the Figure, was found for 2 clusters out of the 20 largest clusters considered for $t_0 =$ January 1st.

6.5.5 Disease sentinels

The study of the initial conditions, and their characterization in terms of the spreading pathways they generate when seeding an outbreak, is crucial for the development of efficient methods to hinder the disease propagation. In particular, the success of control and mitigation measures critically depend on the ability to rapidly detect an outbreak and identify its source. Ideally, a timely detection of the origin of the disease would allow a targeted strategy able to isolate the infected herds and contain the propagation of the infection to other farms. Longer delays between the start of the outbreak and its detection mean larger numbers of infected farms, a more difficult identification of the starting point of the spreading, and therefore a more difficult identification of the propagation pathways, that is needed to prevent further spread to unaffected farms, overall leading to increasingly expensive containment measures. The high temporal variability and the complex nature of the network of displacements makes the identification of the possible origin of the outbreak, following the detection of an infected node, an *a priori* particularly difficult task. However, the very same variable and heterogeneous nature of the network allows for the identification of clusters of seeds leading to similar invasion paths, as discussed in the previous subsections. Here we explore whether the information on the seeds' clusters may provide additional knowledge to enhance surveillance and help the inference of the origin of a disease, once an epidemic is unfolding on the network.

Let us consider a partition $\mathcal{P}(t_0)$ of clusters of initial conditions. If we explore all paths of infections, we can measure the number of times that any node in the network is reached by the epidemic, and break down this number according to the seed cluster originating the epidemic. We can then associate to each node k, reached by the disease n_k times, a vector $\vec{\pi}(k)$ whose components $\pi_j(k)$ represent the probability of being infected by a seeder belonging to the cluster j. If k is reached each of the n_k times by invasion paths rooted in premises belonging to the

same cluster m, the vector has components $\pi_m = 1$ and $\pi_{j\neq m} = 0$. On the contrary, for a node k infected by epidemics originated in farms belonging to a different cluster each of the n_k times, the vector elements assume the values $\pi_j = 1/n_k$. In the case an epidemic is detected at node k by the surveillance system, the vector $\vec{\pi}(k)$ encodes valuable information for the identification of the possible initial condition leading to that epidemic. In particular, it is possible to define an uncertainty $\xi(k)$ in the identification of the seeding cluster, by using an entropy-like function defined as $\xi(k) = -(\log n_k)^{-1} \sum_j \pi_j \log \pi_j$. With reference to the two examples reported before, the uncertainty in the identification of the seeding cluster would be equal to 0 when k is infected by a node in the seed cluster m all the times, and equal to 1 if k is infected each time by a different cluster. Figure 6.22A shows the cumulative distribution of the uncertainty ξ. The number of times n_k that a node is infected may strongly vary from one node to the next; in particular, many nodes are in fact infected just once ($n_k = 1$) in very small outbreaks, yielding trivially high $\xi(k)$ values. We thus focus on farms that have been infected at least 10 times. Interestingly, even with this restriction, the seeder uncertainty is less than 40% for almost 70% of the infected nodes, meaning that most nodes reached by the infection are able to provide valuable insights about the origin of the disease in terms of the identification of the cluster from which the spreading originated. As a result, information about the invasion paths and the epidemic timing is also obtained, following the findings of Figure 6.21.

The uncertainty $\xi(k)$ on the identification of the cluster of initial conditions infecting the node k and the number of times n_k the node k is reached by the epidemic clearly depend on the time t_0 of the start of the epidemic. In the following, we explore the variation of these two quantities for all nodes of the network when we consider epidemics starting at time t_0 = January 1st $+7w$ with $w = 0, 1, 2, 3, ..., 8$, i.e. spanning an 8-weeks interval from January 1st. In Figure 6.22B we represent each farm k as a point with coordinates $(n_k, \xi(k))$ in the two-dimensional $n - \xi$ phase space, for t_0= January 1st. A variety of different behaviors is obtained, as expected given the large variability of the network. Large fluctuations of the number of times a node is infected are observed, as a node with a large n_k (i.e., often reached by the disease) for an initial time t_0 may be rarely reached if the outbreak starts later, given the change in the network of displacements, or may even disappear from the plot if it is not infected for a given explored initial time (i.e. it has $n_k = 0$). Similarly, also the values of the uncertainty in the identification of the seeding cluster can strongly fluctuate. In the surveillance perspective, we are mainly interested in the nodes that are infected a large number of times (i.e. are likely reached by the epidemic, given any temporal and geographical initial conditions) and for which we have a low uncertainty in the

		n_s							
		30	40	50	60	70	80	90	100
ξ_s	0.50	42	40	37	34	31	29	25	25
	0.45	32	31	28	25	22	20	18	18
	0.40	15	14	11	10	9	7	5	5
	0.35	9	8	6	6	5	4	4	4

Table 6.3: Deterministic simulations. Number of sentinels for different thresholds in the $n - \xi$ space.

identification of the seeding cluster, providing important insights into the previous and future spreading patterns. We define these farms as *sentinel nodes* by imposing that they are infected at least n_s times and are characterized by an uncertainty at most equal to ξ_s for all initial conditions. Their trajectories in the $n - \xi$ phase space for varying t_0 are shown in Figure 6.22B, where we have imposed $n_s = 30$ and $\xi_s = 0.4$. The choice of the (n_s, ξ_s) threshold values depends on the resources available to monitor these sentinels: smaller α_s and larger ξ_s lead to a larger number of sentinels.

In Table 6.3, we report the number of sentinels obtained for different choices of the threshold values. It is also possible to be less conservative and enlarge the group of possible proxies for an efficient detection of an infectious disease by including farms with discontinuous trajectories that have $n_k = 0$ for one value of the starting time but have $n_k \geq n_s$ and $\xi(k) \leq \xi_s$ for the other starting times. By relaxing these constraints, it is possible to build a hierarchy of disease sentinels with different levels of reliability, and specific to the available surveillance resources.

The interest of the definition of sentinel nodes in the perspective of a surveillance system is quantified further in the panels C and D of Figure 6.22. Given a set of sentinels, we measure the fraction of detected outbreaks as a function of the outbreak final size, where an outbreak is considered detected if it infects at least a sentinel farm. Figure 6.22C shows that sentinels are not good indicators for the presence of small outbreaks (i.e. corresponding to sizes smaller than 5-10 infected farms), as expected, however a surveillance system based on only 15 sentinel nodes (out of a total number of more than $170,000$ premises) would detect more than 55% of the outbreaks with final size at least 10 and, if the number of sentinels is increased to 32, the fraction of outbreaks detected would be more than 75%. Finally, it is also important to consider that the information provided by the sentinel farms is meaningful as long as the detection occurs rather early in the outbreak evolution. Therefore in Figure 6.22D we evaluate the rapidity of the detection by showing the infection time of each of the 15 sentinel farms (obtained with

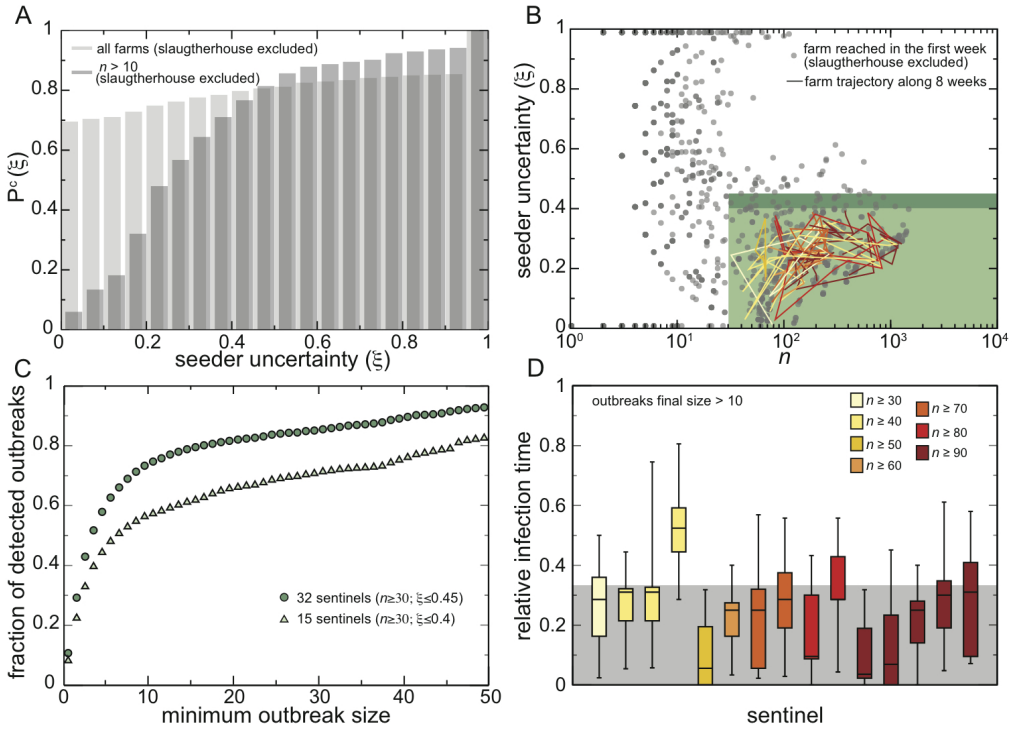

Figure 6.22: Uncertainty in the identification of the initial conditions given the detection of the outbreak and sentinel farms. (A) Cumulative probability distribution of the uncertainty ξ in the identification of the seeding cluster, once a given node of the network is infected. Here slaughterhouse are discarded from the analysis, as they cannot spread the disease further to other farms since they are the end points of the livestock movements and usually gather the bovines from different sources. (B) For a set of initial conditions, each infected farm is represented by a dot in the $n - \xi$ phase space, with n being the number of times the farm is reached by the infection, and ξ the uncertainty in the identification of the corresponding seeding cluster. Eight consecutive weeks starting from January 1st are considered as temporal initial conditions. Sentinel nodes are defined as the farms that are often reached by the epidemic (i.e. $n > n_s$) and that have a low degree of uncertainty in the identification of the seeding cluster that led to the outbreak (i.e. $\xi < \xi_s$). In the plot we show the trajectory in the phase space of the 15 sentinels obtained by imposing $n_s = 30$ and $\xi_s = 0.4$. (C) Fraction of detected outbreaks as a function of the minimum outbreak size of the epidemic, where an outbreak is considered detected if one of the sentinel has been reached by the infection. Two sets of sentinel farms are considered, 15 and 32 sentinels, having both $n_s = 30$ but a different threshold on the uncertainty, i.e. $\xi_s = 0.4$ and $\xi_s = 0.45$, respectively. (D) Boxplot of the time of infection of the 15 sentinels relative to the full duration of the outbreak, considering the detected outbreaks with final size greater than 10. Each box is colored according to the number of times that the sentinel has been infected and a grey shaded area indicates the 33% of the relative infection time.

$n_s = 30$ and $\xi_s = 0.4$) relative to the full outbreak duration, for outbreaks with size larger than 10. Interestingly, almost all sentinels are able to detect most outbreaks within the first third of the outbreak duration.

6.5.6 Generalization to the stochastic case

The high dimensionality of the phase space of the initial conditions has led us to prefer a deterministic computational approach for the presentation of the proposed methodology and main results. We show however in the following how our approach can be extended to stochastic simulations which take into account the intrinsic stochasticity of epidemic propagation phenomena. Results similar to the deterministic case are recovered. In the stochastic simulations, each spreading event is a probabilistic process which occurs with a probability of infection βdt per time interval dt. We keep for simplicity a deterministic recovery process. Moreover, while the weights of the links are irrelevant in the case of a deterministic spreading, they need to be taken into account in a stochastic modeling. In this context, two definitions of weights can be used. We denote by w_{ij}^H the number of herds of cattle and by w_{ij}^B the number of bovines displaced from i to j in the time window Δt. At most one herd is displaced each day, so that w_{ij}^H is at most equal to Δt. Each weight definition leads to a different definition of the probability of infection of a susceptible node by a neighboring infectious node, with different underlying assumptions. A first possibility is to define the rate of infection as $P(S_i + I_j \rightarrow I_i + I_j) = \beta * (w_{ij}^H)/\Delta t$: for $\Delta t = 1$, this means that an infectious node infects a neighboring node to whom it sends a herd with probability β. On the other hand, one can assume that the spreading power is proportional to the number of displaced bovines during each time window. Since the weights w_{ij}^B are broadly distributed (297), we choose to model the probability of infection by a function that saturates to 1 at large values of the weight, namely $P'(S_i + I_j \rightarrow I_i + I_j) = 1 - exp(-\beta' w_{ij}^B/\Delta t)$. We use the finest time scale $\Delta t = 1$ day for the stochastic simulations. In this case, the number of herds displaced between two nodes at each time step is either 0 or 1. In order to compare the two possible assumptions underlying the stochastic simulations, we use values of β and β' such that the probabilities P and P' are equal on average. This condition is satisfied if $\beta' = ln(1-\beta)/\langle w_{ij}^B \rangle$. We explore two values of β: high transmission rate ($\beta = 0.9$, corresponding to $\beta' = 0.63$) and intermediate transmission rate ($\beta = 0.5$, corresponding to $\beta' = 0.19$). Once the transmission probabilities are defined and the transmission rate fixed, each stochastic simulation produces an invasion path. The union of all the paths yields a risk probability associated to each node, given

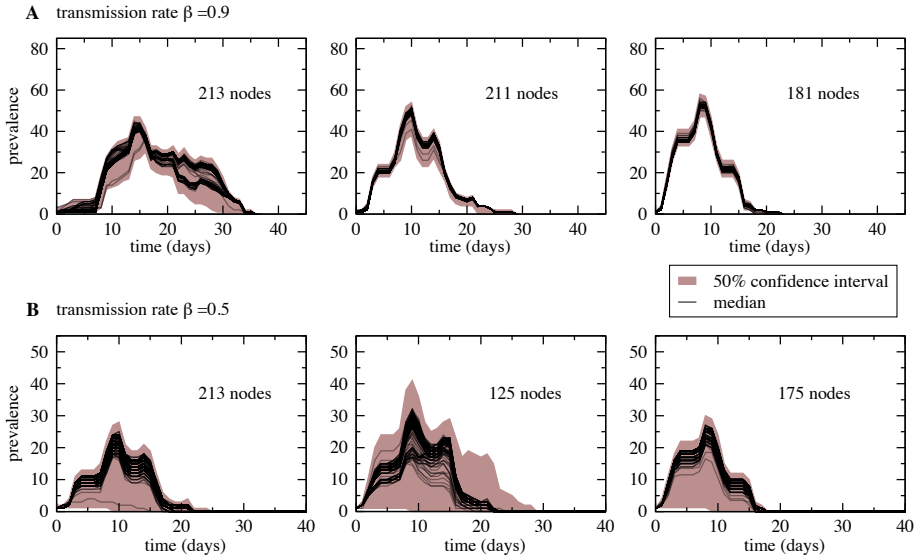

Figure 6.23: Prevalence curves of stochastic spreading starting from nodes belonging to various clusters. For each seed 500 stochastic simulations have been performed and the medians (black curves) and 50% confidence intervals are shown. In order to give an estimate of the fluctuations we evaluated the 50% confidence interval for each curve and we plot the maximum of the upper bounds and the minimum of the lower bounds (brown shaded area). Higher transmission rate (A) leads to lower fluctuations, while for the intermediate transmission rate (B), the 50% confidence intervals are quite broad. It is worth to stress that hundreds of stochastic curves with different seeding nodes are compared, so some fluctuations are expected; nevertheless the clustering procedure is still able to capture the similarity in the overall behavior (shape, peak time, duration) of the prevalence curves.

by the fraction of runs in which the node has been infected. We generalize the construction of the ICSN and of the clusters defined in the previous section for deterministic spreading as follows. For each seed x, we define a vector r whose element r_i is given by the fraction of runs starting in x for which i was infected, and the set ν of nodes i such that $r_i > 0$ (i.e., the set of nodes which were reached at least once by a spreading issued from x). For each pair of seeds x_1 and x_2, we build in this way the two vectors r_1 and r_2 and the sets of nodes ν_1 and ν_2, and we consider the similarity $\Omega_{12} = \sum_i (1 - |r_{1,i} - r_{2,i}|^2)/|\nu_1 \bigcup \nu_2|$. In the case of a deterministic spreading, we recover the definition of the overlap Θ as the elements of the vectors r_i are 0 or 1.

The cluster detection method described previously can now be applied using this measure by retaining in the ICSN only the links with similarity Ω larger than a certain threshold.

We show in Figure 6.23 the prevalence curves for seeds belonging to various clusters. As in

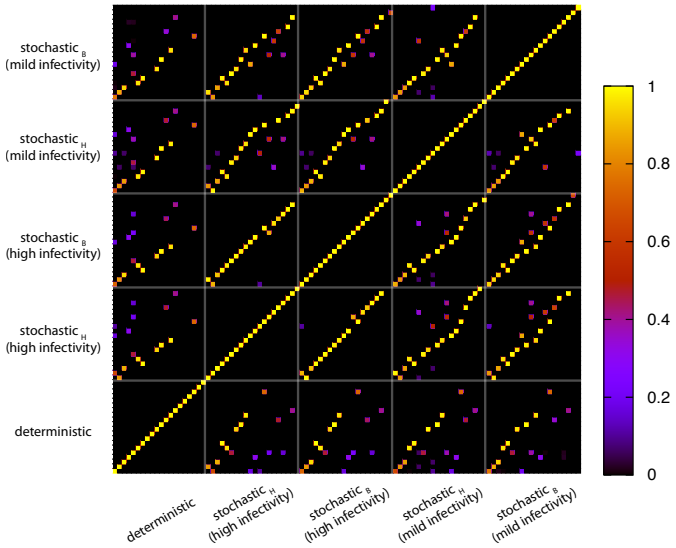

Figure 6.24: Jaccard indices of the 20 largest clusters obtained with different simulation procedures. We refer with stochastic$_H$ and stochastic$_B$ to simulations with transmission probabilities respectively determined by w_{ij}^H and w_{ij}^B. Each point represents the intersection of different clusters obtained with different transmission probability definitions and transmission rates and is color-coded according to the value of the Jaccard index computed as the number of common nodes divided by the number of nodes in the union of the clusters. The violet-to-yellow color scale indicates how much the clusters obtained in different simulation procedures have in common.

		n_s							
		30	40	50	60	70	80	90	100
ξ_s	0.45	39	34	32	27	24	21	20	16
	0.40	27	24	22	18	16	14	13	10
	0.35	15	13	12	10	8	6	5	3
	0.30	4	4	4	2	2	1	1	0

Table 6.4: Stochastic simulations. Number of sentinels for different thresholds in the $n - \xi$ space.

the deterministic case, spreading phenomena originated in nodes of the same cluster show very similar temporal patterns. The cluster detection procedure leads thus to an efficient grouping of the potential seeds of an epidemics not only for deterministic spreading but also for more realistic stochastic simulations.

We also evaluate and show in Figure 6.24 the overlap between the clusters obtained with the various types of stochastic simulations and the deterministic ones. Strikingly, many clusters are very stable when the type of simulation and the infectivity parameter are changed.

Once the nodes are grouped in different clusters, the sentinel identification is a rather straightforward procedure. Using the same definition of seeder uncertainty described in the main paper, it is thus possible to identify sentinel nodes starting from stochastic disease spreading simulations. It is worth to notice that for each seeding node we simulate 100 stochastic runs, so that the number of times n_k that a node k has been reached by the disease is, potentially, much larger than the deterministic case. In order to make comparable the stochastic and the deterministic scenarios we rescale n in Figure 6.25 and table 6.4 of a factor 100. In Figure 6.25 we show the results obtained from stochastic simulations, similarly to Figure 6.22. Surprisingly, the stochastic simulations lead to a lower uncertainty than the deterministic case. This counter-intuitive behavior can be attributed to the fact that including some heterogeneities due to the links' weight the less probable invasion paths contribute very little in the seeder uncertainty evaluation, naturally reducing the noise of the measure as already shown for the spreading on the airport network in (110). Since the choice of ξ_s and n_s are arbitrary, we report in table 6.4 the number of sentinels identified with different values.

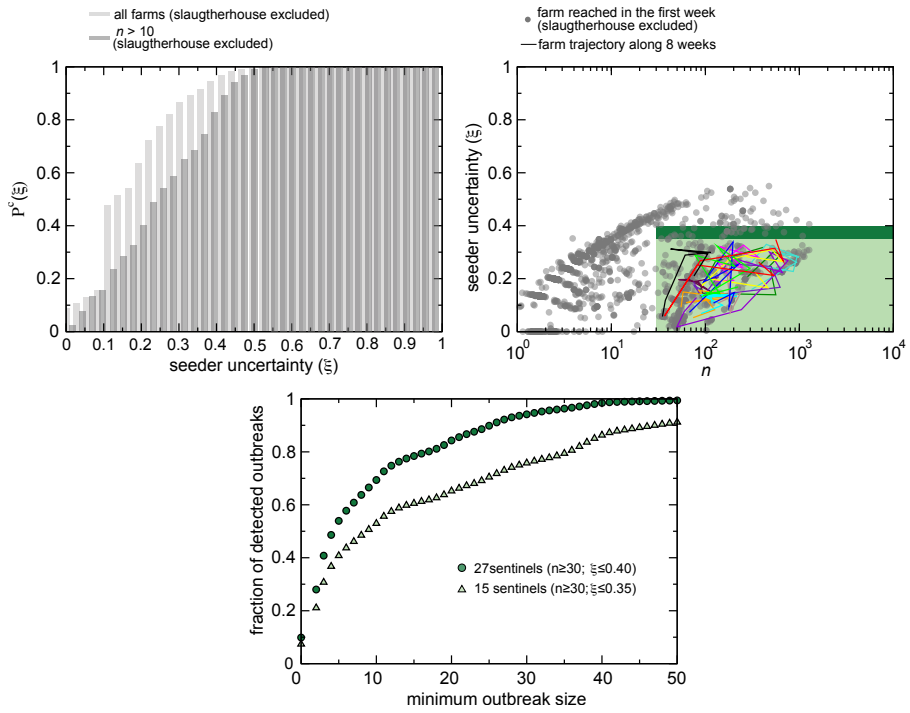

Figure 6.25: Identification of the initial conditions, and corresponding uncertainty similarly to Figure 6.22, for the stochastic simulations. (A) Cumulative probability distribution of the uncertainty ξ in the identification of the seeding cluster, once a given node of the network has been detected as infected. (B) For a set of initial conditions, each infected farm is represented by a dot in the $n - \xi$ phase space, with n being the number of times the farm is reached by an infection, and ξ the uncertainty in the identification of the corresponding seeding cluster. (C) Fraction of detected outbreaks as a function of the minimum outbreak size of the epidemic. Two sets of sentinel farms are considered, of 15 and 27 sentinels, corresponding respectively to ($n_s = 30$, $\xi_s = 0.35$) and ($n_s = 30$, $\xi_s = 0.40$).

6.6 Conclusions

In this chapter, we have presented a full analysis of a dynamical networked system, going beyond static and simple approximations and taking fully into account the temporal dimension of the dataset, using the Italian data of cattle movements of 2007 as a prototypical example. Starting from detailed data at the individual level at a daily resolution and covering a whole year, we have constructed aggregated networks on different timescales to characterize the system's behavior on a variety of timescales, exposing the coexistence of stationary statistical distributions and strong microscopic dynamics at all time and spatial scales. We have shown how this dynamics affects not only global quantities (such as the number of connected nodes), but also the nodes' and links' properties at a very local level, and in relation with the rest of the system. In particular, the centrality of a node fluctuates strongly in time, thus preventing a straightforward static assessment of the spreading potential of premises that could be used for the definition of prevention and control measures. The network's dynamics also hinders the definition of a stationary backbone for the system structure and function, as a subset of the most important links (and weights) that are stable over time. We found indeed that the nodes and links forming the backbone strongly vary depending on the time window considered, and that the memory of the backbone rapidly fades away from one snapshot to the successive ones. We have put forward a definition of dynamical motifs, formed by sequences of links that allow causal propagation, and illustrated how this definition can unveil the existence of an intrinsic time arrow in the dataset. The number of motifs of various lengths is indeed strongly different in the real dataset and in a time-reversed version; moreover such definition can be easily extended to focus on a variety of timescales of interest for the study of different dynamical processes.

The strong temporal fluctuations of the system have important implications for the dynamical phenomena occurring on the system. Evaluating the information available at a given time step, to devise containment strategies against an epidemic spreading on the system, would indeed lead to inefficient measures if applied at other times. Through simulations on the fully dynamic network, where daily bovine movements are explicitly captured, we have studied the role of the initial conditions (i.e. seeds) in shaping the propagation process. Clusters of seeds emerge that lead to similar spreading patterns in terms of infected farms, and are also characterized by similar epidemic profiles and peak times. These clusters cannot be trivially identified from structural or geographical considerations. The proposed clustering method can be used in order to define rapid and efficient containment strategies, targeting farms that are at high risk of being infected

and further spread the disease. Although the displacement network is characterized by a large temporal variability, intrinsically altering the centrality role of nodes from a given observation time to another, it is however possible to identify sentinel nodes representing premises which are often reached by the disease and, when detected as infected, are able to provide valuable information on the seeding farms of the outbreak. Remarkably, the complex aspects of the dynamical network and their interplay with the disease dynamics lead to the emergence of a very small number of sentinel premises, with respect to the total number present in the system, that may be efficiently used for disease prevention and control. A hierarchical classification of sentinels can be provided by tuning the constraints imposed for their definition, thus leading to different levels of surveillance. Applications to specific diseases, where the timescale of the epidemic is set by the parameters describing the specific disease etiology, can be performed to tune this framework to particular cases. These findings clearly depend on the full knowledge of the displacement dataset, and can thus be obtained as a priori information during a non-emergency period to help orienting control strategies, as it is commonly done with the static analysis of the contact network structure. However, it is not obvious how to exploit the information provided by the sentinel nodes in order to explicitly track back the disease pathway and identify the origin of the epidemic. Future research will focus on a possible reverse-engineering process aimed at identifying a list of possible seeder nodes once that an outbreak has been detected by a sentinel. In addition, analysis of successive years of movements data, uncovering possible recurrent patterns and seasonal behaviors, may contribute to make this framework a general tool to be used in real-time emergencies.

7

Conclusions and perspectives

In this thesis we have presented theoretical and computational frameworks for the investigation of disease spreading phenomena in real settings. Following the findings and the results achieved by recent studies on complex networked systems and recognizing in the interconnectivity among host populations the key ingredient for the dissemination of infectious diseases, we investigated large-scale empirical datasets providing crucial information about the underlying host mobility structure in emergent epidemics. The first part of this thesis was devoted to present the Global Epidemic and Mobility model (GLEaM) and its application to the past 2009 H1N1 pandemic. We have described the disease parameters estimation through a human mobility based Monte Carlo likelihood method, and we have presented the real-time projections and results that we had achieved well before the epidemic's peak. Then we have focused on possible intervention strategies highlighting the limited efficacy of the vaccination campaign given the early occurrence of the pandemic's peak, and we have discussed both theoretically and numerically the scarce effectiveness of travel restriction policies for containing the new virus at the source. The 2009 pandemic influenza revealed the vulnerability of our interconnected world to the spreading of new pathogens and demonstrated the need of computational models to address public health issues related to intervention policies aimed at monitoring, controlling and preventing large scale outbreaks. In this perspective, we think that our work provided an important contribution to the multidisciplinary research in the direction of a data-driven real-time epidemic forecasting. We are aware that lots of work has to be done in order to include a more structured epidemic model able to integrate a higher level of realism. In particular, it would be interesting to include social structures and behavioral changes in the synthetic population, along with an even more accurate mobility network able to capture the seasonal nature of commercial air flights and possible long

term migrations. On the other hand, the more intricate and detailed a model is, the wider are the assumptions and the larger is the amount of data needed to estimate parameters. It is thus necessary to assess how the present details are important and drive the forecasts precision, and how our multi-scale GLEaM model is susceptible to small perturbations of the initial conditions. The predictive power and the fundamental limits in epidemic evolution predictability with computational modeling represent the main open issues that modelers and policy makers have to face in the next years.

In the last part of this thesis, we have tackled the new and challenging problems related to the analysis of dynamical networks and the interplay of the system dynamics with the dynamical processes unfolding on top of them. The highly detailed dataset of livestock movements represented a unique case study to test the robustness, the reliability and the limits of a static approach and to devise new mathematical tools aimed at deeper investigations. We have performed an extensive and systematic longitudinal analysis unveiling the coexistence of the stationarity of global distributions along with a microscopic blinking topology. We quantitatively analyzed the temporal fluctuations and we put forward a novel definition of dynamical motifs able to uncover temporal correlations in the network dynamic. It is worth to notice that the study of dynamical processes on a static representation of the system may lead to limited and inaccurate results, and a new perspective is needed to handle processes entangled with the network dynamics. Beyond the system characterization, we have thus studied the dissemination of an emerging infectious livestock disease and we have proposed a new method to cluster nodes with similar spreading potential properties. We have further indicated a novel procedure to identify sentinel nodes able to largely improve the surveillance system efficacy. Future developments of this research will include a more structured epidemic model and an assessment of the proposed method with real epidemiological data. Furthermore, the reverse-engineering process able to reconstruct the spreading pathways given the detection of an outbreak by means of a disease sentinel node is still under study. We dare to hope that our work could contribute to pose the basis of a taylored approach to the analysis of evolving networked systems, and we believe that the increasing availability of different longitudinal datasets will provide vital insights of emerging phenomena in realistic settings.

Acknowledgements

I owe an immense debt of gratitude to Vittoria Colizza for giving me the opportunity to work on such stimulating topics in one of the most pleasant place for doing Science, namely the I.S.I. Foundation. Here I had the chance to interact, discuss, work, play football and have "aperitivi" with nice people and top scientists in their field.

I give my greatest thanks to Alain Barrat that has been the ideal thesis supervisor. His patient encouragement, insightful criticisms, and wise advices aided the writing of this thesis in innumerable ways. I am grateful to Vittoria, Alain and Alessandro Vespignani for their sincere and spontaneous frame of mind and for bringing me to the exciting side of research and for teaching me how to do (hopefully good) Science.

I should acknowledge many people for these three beautiful, dense and exciting years of my PhD training. All my colleagues in Turin and Marseille, those who I have met during the first months and then left, and those who have arrived in the last period, contributed to make the office a friendly and stimulating workplace. In particular, I warmly thank Chiara, Michele, Daniela, Ciro, Corrado, Duygu, Juliette, Lanci and Santo for being much more than exceptional colleagues. To all these friends I am deeply grateful for their advices and inspiring discussion about life, movies, music, restaurant (or food and beverage in general), sports, politics, nerdy stuff, and, of course, complex systems.

I want to thank the Bloomingtonians friends, Nicola, Bruno, Hao, Fabio and Duygu (again): even though they have moved, their names are indissolubly related to that college town where they have warmly welcomed me and supported during my stay.

The European Research Council Ideas contract n.ERC-2007-Stg204863 (EPIFOR) that partially funded this work is gratefully acknowledged.

Finally, special thanks goes to my dearest friends and my beloved family. I believe that the strength of their affection and emotional support along with their teaching and encouragements goes beyond every obstacle and difficulty. Thanks to Eleonora, my accomplice, my friend, my wife.

References

[1] U. Alon. Biological networks: the tinkerer as an engineer. *Science*, **301**:1866–1867, 2003. 1, 10

[2] A.-L. Barabási and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nat. Rev. Gen.*, **5**:101–113, 2004. 1, 10

[3] H. Jeong, S. Mason, A.-L. Barábasi, and Z.N. Oltvai. Lethality and centrality in protein networks. *Nature*, **411**:41–42, 2001. 1, 9, 10

[4] J.A. Dunne, R.J. Williams, and N.D. Martinez. Food-web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci.*, 99:12917–12922, 2002. 1, 10

[5] H. Ebel, L.I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, **66**:035103, 2002. 1, 9

[6] M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, **66**:035101, 2002c. 1, 9

[7] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998. 1, 9, 10, 12

[8] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, **286**:509–512, 1999. 1, 9, 13, 89

[9] J.J. Ramasco, S.N. Dorogovtsev, and R. Pastor-Satorras. Self-organization of collaboration networks. *Phys. Rev. E*, **70**:036106, 2004. 1, 9

[10] M.E.J. Newman. Scientific collaboration networks. I. Network contruction and fundamental results. *Phys. Rev. E*, **64**:016131, 2001. 1, 9

[11] M.E.J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks and centrality. *Phys. Rev. E*, **64**:016132, 2001. 1, 9

[12] M.E.J. Newman. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci.*, **98**:404–409, 2001. 1, 9

[13] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Staneeley. Classes of small-world networks. *Proceeding of the National Academy of Science (USA)*, **97**:11149–11152, 2000. 1, 10, 87

[14] A. De Montis, M. Barthélemy, A. Chessa, and A. Vespignani. The structure of inter-urban traffic: A weighted network analysis. *Env. Planning Journal B*, 2006. 1, 9, 10, 76, 78, 87, 93

[15] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architeture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, **101**:3747–3752, 2004a. 1, 9, 10, 32, 76, 78, 91, 93, 97, 104

[16] R. Guimerá and L.A.N. Amaral. Modeling the world-wide airport network. *Eur. Phys. J.B.*, **38**:381–385, 2004. 1, 9, 10, 93

[17] G. Bonanno, G. Caldarelli, F. Lillo, and R.N. Mantegna. Topology of correlation based minimal spanning trees and model markets. *Phys. Rev. E*, 68:046130, 2003. 1

[18] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccicché, N. Vandewalle, and R.N. Mantegna. Networks of equities in financial markets. *Eur. Phys. J. B*, 38:363, 2004. 1

[19] D Garlaschelli, S. Battiston, S. Castri, M. Servedio, and G. Caldarelli. The scale free topology of market investments. *Physica A.*, **350**:491–499, 2005. 1

[20] C.A. Hidalgo and R. Hausmann. The building blocks of economic complexity. *Proc. Natl. Acad. Sci. USA*, 106(26):10570–10575, 2009. 1

[21] M.C. González, C.A. Hidalgo, and A.L. Barabási. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008. 1, 9, 91, 92, 93, 117

[22] D. Brockmann, L. Hufnagel, and L. Geisel. The scaling laws of human travel. *Nature*, **439**:462–465, 2006. 1, 9, 92, 93

[23] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabó, D. Lazer, et al. Structure and tie strngths in mobile communication networks. *Proc. Natl Acad. Sci.*, 104:7332–7336, 2007. 1

[24] B. Goncalves, N. Perra, and A. Vespignani. Modeling users' activity on twitter networks: Validation of dunbar's number. *PLoS ONE*, 6(8): e22656, 2011. 1

[25] M. Granovetter. Strength of weak ties. *American Journal of Sociology*, **78**:1360–1380, 1973. 1

[26] R.I.M. Dunbar. Neocortex size as a constraint on group size in primates. *J. Human Evo.*, 22:469, 1992. 1

[27] R. Pastor-Satorras and A. Vespignani. *Evolution and Structure of Internet: A Statistical Physics Approach*. Cambridge University Press., 2004. 1, 6, 10, 23, 89, 95

[28] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world wide web. *Nature*, **401**:130–131, 1999. 1, 10

[29] A. Vespignani. Predicting the behavior of techno-social systems. *Science*, 325:425–428, 2009. 2

[30] Bergé, C. *Graphs and Hypergraphs*. North-Holland, 1976. 6

[31] Chartrand, G. and Lesniak, L. *Graphs and Digraphs*. Wadsworth and Brooks/Cole, 1986. 6

[32] Bollobás, B. *Random Graphs*. Cambridge studies in advanced mathematics, 1985. 6

[33] Clark, J. and Holton, D.A. *A First Look at Graph Theory*. World Scientific, 1991. 6

[34] Harary, F. *Graph Theory*. Perseus, 1995. 6

[35] West, G.B. *Introduction to Graph Theory*. Prentice Hall, 1996. 6

[36] Bollobás, B. *Modern Graph Theory*. Springer-Verlag, 1998. 6

[37] Barrat, A. and Barthélemy, M. and Vespignani, A. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008. 6, 7, 89, 90, 95

[38] Caldarelli, G. *Scale-Free Networks Complex Webs in Nature and Technology*. Oxford Finance Series, 2007. 6, 7, 89, 95

[39] S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of Networks From Biological Nets to the Internet and the WWW.* Oxford University Press., 2003. 6, 7, 89, 95

[40] Newman, M.E.J. *Networks, an Introduction.* Oxford University Press, 2010. 6, 7

[41] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press, 2010. 6

[42] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, **89**:208701, 2002a. 8

[43] G. others Chowell. 9, 76, 78, 93

[44] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184, 2004. 9, 45

[45] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *PNAS*, 103:2015, 2006. 9, 27

[46] P. Kaluza, A. Kolzsch, M.T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *J R Soc Interface*, 10.1098/rsif.2009.0495, 2010. 9, 93

[47] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabó, M. Argollo de Menezes, et al. Analysis of a large-scale weighted network of one-to-one human communication. *New J Phys*, 9:179, 2007. 9

[48] C. Song, Z. Qu, N. Blumm, and A.L. Barabási. Limits of predictability in human mobility. *Science*, 327:1018–1021, 2010. 9, 91, 92

[49] A. Gautreau, A. Barrat, and M. Barthélemy. Microdynamics in stationary complex networks. *Proc Natl Acad Sci USA*, 22:847–8852, 2009. 9, 102, 105, 106, 108, 109

[50] A.L. Barabási. Scale-free networks: a decade and beyond. *Science*, 325:412–413, 2009. 9

[51] Moreno, J.L. *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodram.* Beacon House, 1934. 9

[52] A.-L. Barabási and R. Albert. *Nature*, **286**:509, 1999. 9

[53] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. *Physical Review E*, **64**:026118, 2001. 9

[54] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Computer Communications Review*, **29**:251–262, 1999. 10

[55] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. The origin of power laws in Internet topologies revisited. *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, 2002. 10

[56] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph Structure in the Web. *Computer Networks*, **33**:309–320, 2000. 10

[57] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, and Y. Aberg. The web of human sexual contacts. *Nature*, **411**:907–908, 2001. 10, 87

[58] P. Erdös and A Rényi. *Publ. Math.*, **6**:290, 1959. 10

[59] P. Erdös and A Rényi. *Publ. Math. Inst. Hung. Acad. Sci.*, **5**:17, 1960. 10

[60] P. Erdös and A Rényi. *Bull. Inst. Int. Stat.*, **38**:343, 1961. 10

[61] F. Chung and .H Lu. *Advances Applied Mathematics*, **26**:257, 2001. 11

[62] A. Barrat and M. Weigt. *Eur. Phys. J. B*, **13**:547, 2000. 13

[63] A.-L. Barabási, R. Albert, and H. Jeong. *Physica A*, **272**:173, 1999. 13

[64] B. Bollobás and O. Riordan. The diameter of scale-free random graphs. 2002. 13

[65] K. Klemm and V.M. Eguiluz. Growing scale-free networks with small-world behavior. *Physical Review E*, 65, 2002. 13

[66] G. Szabó, M. Alava, and J. Kertész. Structural transitions in scal-free networks. *Physical Review E*, **67**, 2003. 13

[67] A. Barrat and R. Pastor-Satorras. Rate equation approach for correlation in growing network models. *Physical Review E*, **71**, 2005. 13

[68] L.E.C. Rocha, F. Liljeros, and P. Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comp Biol*, 7:e1001109, 2011. 15

[69] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, et al. Simulation of an sir infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9:87, 2011. 15

[70] P. Holme and J. Saramäki. Temporal networks. *arXiv:1108.1780*, 2011. 15, 113

[71] Anderson, R.M. and May, R.M. *Infectious Diseases in Humans*. Oxford Univ. Press, 1992. 18, 20, 23, 27, 46, 81, 119

[72] Bailey, N.T. *The mathematical theory of infectious diseases*. Griffin, 1975. 18, 22

[73] Keeling, M.J. and Rohani, P. *Modeling infectious diseases in humans and animals*. Pronceton Univeristy Press, 2008. 18, 22

[74] Vynnycky, E. and White, R. *An Introduction to Infectious Disease Modelling*. Oxford Univeristy Press, 2010. 18

[75] Daley, D.J. and Gani, j. *Epidemic Modelling*. Cambridge Univeristy Press, 1999. 18

[76] O. Diekmann, A.J.P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations. *J. Math. Bio.*, 28:1432, 1990. 18, 46

[77] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001. 23, 90

[78] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J.J. Ramasco, and Vespignani A. Multiscale mobility networks and the large scale spreading of infectious diseases. *Proc. Natl Acad. Sci. USA*, page 106:21484, 2009. 23, 26, 27, 32, 35, 43, 60, 63, 66, 91, 92, 93, 122

[79] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Epidemic spreading in complex networks with degree correlations. *Lect. Notes Phys.*, 625:127–147, 2004. 23

[80] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B*, 26:521–529, 2004. 23

[81] K. Klemm and V.M. Eguiluz. Highly clustered scale-free networks. *Physical Review E*, 65, 2002. 23

[82] V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature Phys.*, 3:276, 2007. 25, 76, 79, 90

[83] R.M. Anderson and R.M. May. Spatial, temporal and genetic heterogeneity in hosts populations and the design of immunization programs. *IMA J. Math. Appl. Med. Biol.*, 1:233–266, 1984. 26

[84] B.M. Bolker and Grenfell. B.T. Space persistence and dynamics of measles epidemics. *Phil. Trans. Biol. Sci.*, 348:309–320, 1995. 26, 27

[85] A.L. Lloyd and R.M. May. Spatial heterogeneity in epidemic models. *J. Theor. Biol.*, 179:1–11, 1996. 26, 27

[86] N.M. Ferguson, M.J. Keeling, W.J. Edmunds, et al. Planning for smallpox outbreaks. *Nature*, 425:681–685, 2003. 26

[87] D.J. Watts, R. Muhamad, D.C. Medina, and P.S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulations model. *Proc. Natl. Acad. Sci.*, 102:11157–11162, 2005. 26

[88] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. *Bull Math Biol*, **68**:1893–1921, 2006. 26, 32

[89] C. Viboud, O.N. Bjornstad, D.L. Smith, L. Simonsen, M.A. Miller, and B.T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312:447, 2006. 26

[90] O.V. Baroyan, L.A. Genchikov, L.A. Rvachev, and V.A. Shashkov. An attempt at large-scale influenza epidemic modelling by means of a computer. *Bull. Int. Epidemiol. Assoc.*, 18:22–31, 1969. 27

[91] L.A. Rvachev and I.M. Longini. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75:3–22, 1985. 27, 52

[92] I.M. Longini. A mathematical model for predicting the geographic spread of new infectious agents. *Math. Biosci*, 90:367–383, 1988. 27

[93] R.F. Grais, H.J. Ellis, and G.E. Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur. J. Epidemiol.*, 18:1065â1072, 2003. 27

[94] R.F. Grais, J.H. Ellis, A. Kress, and G.E. Glass. Modeling the spread of annual influenza epidemics in the u.s.: The potential role of air travel. *Health Care Manag Sci*, 7:127, 2004. 27

[95] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci.*, 101:15124, 2004. 27

[96] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J.J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani. Seasonal transmission potential and activity peaks of the new influenza a(h1n1): a monte carlo likelihood analysis based on human mobility. *BMC Medicine*, **439**:7:45, 2009. 27, 52, 58, 59, 60, 63, 66

[97] D.J.D. Earn, P. Rohani, and B.T. Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proc. Roy. Soc. Lond. B*, 265:7–10, 1998. 27

[98] P. Rohani, D.J.D. Earn, and B.T. Grenfell. Opposite patterns of synchrony in sympatric disease metapopulations. *Science*, 286:968–971, 1999. 27

[99] M.J. Keeling. Metapopulation moments: coupling, stochasticity and persistence. *Journal of Animal Ecology*, 69:725–736, 2000. 27

[100] A.W. Park, S. Gubbins, and C.A. Gilligan. Extinction times for closed epidemics: the effects of host spatial structure. *Ecology Letters*, 5:747–755, 2002. 27

[101] F. Carrat, E. Vergu, N.M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A.J. Valleron. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *Am J Epidemiol*, **167**:775–785, 2008. 27, 46, 50

[102] L. Canini and F. Carrat. Population modeling of influenza a/h1n1 virus kinetics and symptom dynamics. *Jour Virol*, **85**:2764–2770, 2011. 27

[103] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 5(3): e74, 2008. 27

[104] P. Rohani, X. Zhong, and A.A. King. Contact network structure explains the changing epidemiology of pertussis. *Science*, 330:982–985, 2010. 27

[105] J. Shaman and M.A. Kohn. Absolute humidity modulates influenza survival, transmission and seasonality. *Proc Natl Acad Sci USA*, 106:3243–3248, 2009. 27

[106] J. Shaman, V.E. Pitzer, C. Viboud, B.T. Grenfell, and M. Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental united states. *PLoS Biol*, 8:e1000316, 2010. 27

[107] J. Shaman, E. Goldstein, and M. Lipsitch. Absolute humidity and pandemic versus epidemic influenza. *Am J Epidemiol*, **173**:127–135, 2011. 27

[108] Center for international earth science information network (ciesin), columbia university; and centro internacional de agricultura tropical (ciat). the gridded population of the world version 3 (gpwv3): Population grids. palisades, ny: Socioeconomic data and applications center (sedac), columbia university. 30

[109] Center for international earth science information network (ciesin), columbia university; international food policy research institute (ifpri); the world bank; and centro internacional de agricultura tropical (ciat). global rural-urban mapping project (grump), alpha version: Population grids. palisades, ny: Socioeconomic data and applications center (sedac), columbia university. 30

[110] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.*, **103**:2015–2020, 2006. 32, 136

[111] V. Colizza, A. Barrat, M. Barthélemy, A.J. Valleron, and A. Vespignani. Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Med*, **4**, 2007. 32, 45, 50, 60, 61, 71, 76

[112] L. Sattenspiel and K. Dietz. A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences*, **128**:71–91, 1995. 38

[113] M.J. Keeling and P. Rohani. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology Letters*, **5**:20–29, 2002. 39

[114] D. Balcan, B. Goncalves, H. Hu, J.J. Ramasco, V. Colizza, and Vespignani A. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, 1:3:132–145, 2010. 43

[115] N.M. Ferguson, D.A.T. Cummings, C. Fraser, J.C. Cajka, P.C Cooley, and D.S. Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442:448–452, 2006. 45, 60, 61, 66, 71

[116] C. Germann, K. Kadau, I.M. Longini, and C.A. Macken. Mitigation strategies for for pandemic influenza in the united states. *Proc. Nat. Acad. Sci.*, 103:5935–5940, 2006. 45, 60, 61, 66, 71

[117] Who, pandemic (h1n1) 2009 briefing note 3 (revised): Changes in reporting requirements for pandemic (h1n1) 2009 virus infection. 45, 55

[118] C. Fraser et al. Pandemic potential of a strain of influenza a(h1n1): early findings. *Science*, 324:1557–1561, 2009. 45, 46, 47, 50, 58, 59

[119] G. Cruz-Pacheco, L. Duran, L. Esteva, A.A. Minzoni, M. López-Cervantes, P. Panayotaros, A. Ahued, and I. Villaseñor. Modelling of the influenza a(h1n1)v outbreak in mexico city, april-may 2009, with control sanitary measures. *Euro Surveill*, 14:19254, 2009. 45, 47, 60

[120] K. Khan, J. Arino, W. Hu, P. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, A. Chan, and M. Gardam. Spread of a novel influenza a(h1n1) virus via global airline transportation. *N Engl J Med*, 361:212–214, 2009. 45

[121] I.M. Longini, M.E. Halloran, A. Nizam, and Y. Yang. Containing pandemic influenza with antiviral agents. *Am J Epidemiol*, 159:623, 2004. 46, 50, 60, 61, 63, 66

[122] I.M. Longini, A. Nizam, Xu S., K. Ungchusak, W. Hanshaoworakul, D.A.T. Cummings, and M.E. Halloran. Containing pandemic influenza at the source. *Science*, 309:1083–1087, 2005. 46, 50, 60, 61, 63, 66

[123] Brote de infeccion respiratoria aguda en la gloria, municipio de perote, mexico secretaria de salud, mexico. 47

[124] B.S. Cooper, R.J. Pitman, W.J. Edmunds, and N.J. Gay. Delaying the international spread of pandemic influenza. *PLoS Medicine*, 3:e212, 2006. 48, 54, 71, 76

[125] Who wkly epidemiol rec, 2009. 50

[126] Cdc interim guidance for clinicians on identifying and caring for patients with swine-origin influenza a (h1n1) virus infection, 2009. 50

[127] F.S. Dawood, S. Jain, L. Finelli, M.W. Shaw, S. Lindstrom, et al. Novel swine-origin influenza a (h1n1) virus investigation: Emergence of a novel swine-origin influenza a (h1n1) virus in humans. *N Engl J Med*, 360:2605–2615, 2009. 50

[128] M.J. Roberts and J.A.P. Heesterbeek. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J. Math. Bio.*, 55:803–816, 2007. 50

[129] J. Wallinga and M. Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B*, **274**:599–604, 2007. 50

[130] R. Gani, C. H. Hughes, D. Fleming, T. Griffin, J. Medlock, and S. Leach. Potential impact of antiviral drug use during influenza pandemic. *Emer. Inf. Dis.*, 11:1355, 2005. 50, 60, 61, 66

[131] L.R. Elveback, J.P. Fox, E. Ackerman, A. Langworthy, M. Boyd, and L. Gatewood. An influenza simulation model for immunization studies. *Am J Epidemiol*, 103:152–165, 1976. 50

[132] P.Y. Boelle, P. Bernillon, and J.C. Desenclos. A preliminary estimation of the reproduction ratio for new influenza a(h1n1) from the outbreak in mexico, march-april 2009. *EuroSurveilliance*, 14, 2009. 50, 52

[133] H. Nishiura, C. Castillo-Chavez, M. Safan, and G. Chowell. Transmission potential of the new influenza a(h1n1) virus and its age-specificity in japan. *EuroSurveilliance*, 14, 2009. 50, 52

[134] H. Nishiura, N.M. Wilson, and M.G. Baker. Estimating the reproduction number of the novel influenza a virus (h1n1) in a southern hemisphere setting: preliminary estimate in new zealand. *NZ Med J*, 122:1–5, 2009. 50, 52

[135] J. Lessler, N.G. Reich, R. Brookmeyer, T.M. Perl, K.E. Nelson, and D.A. Cummings. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect DisA*, 9:291–300, 2009. 50

[136] Cdc: Briefing on public health investigation of human cases of swine influenza, april 23, 2009, 3:30 p.m. es. 51

[137] Public health agency of canada, cases of h1n1 flu virus in canada, june 10 2009. 51

[138] King's-edgehill school web site, first cases in canada was related to a school trip in mexico. 51

[139] Who, chronology of influenza a(h1n1). 51

[140] El salvador journal,primeros casos de gripe porcina el salvador, may 4 2009. 51

[141] Ministerio de salud pública y asistencia social, 8th official update, may 3, 2009. 51

[142] Abc news, swine flu cases confirmed in scotland, april 28. 51

[143] The scottish government, scottish government news release, april 26. 51

[144] Ministerio de sanidad y political social, official update, april 27 2009. 51

[145] The guardian, spain confirms first swine flu case in europe, april 27 2009. 51

[146] Usa today, cuba confirms its 1st swine flu case, may 12 2009. 51

[147] Global post, costa rica reports first swine flu case, april 28 2009. 51

[148] Ministry of health welfare and sport (netherlands), first victim mexican flu, april 30 2009. 51

[149] Robert koch institut, neue influenza a/h1n1 in deutschland bewertung des bisherigen geschehens. 51

[150] Ministère de la santé et des sports, official update, may 1. 51

[151] El periodico guatemala, ministro de salud confirma primer caso de ah1n1 en guatemala, may 5 2009. 51

[152] Ministerio de salud pública de guatemala, official update, may 5 2009. 51

[153] Ministerio de la protecciìon social república de colombia, official update, may 3 2009. 51

[154] C.E. Mills, J.M. Robins, and M. Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 432:904–906, 2004. 52

[155] N. Wilson and M.G. Baker. The emerging influenza pandemic: estimating the case fatality ratio. *Euro Surveilliance*, 14:26, 2009. 56

[156] T. Garske et al. Assessing the severity of the novel a/h1n1 pandemic. *BMJ*, 339:b2840, 2009. 56

[157] M. Lipsitch, M. La jous, J.J. OHagan, T. Cohen, and J.C. Miller. *PLoS ONE*, 4:e6895, 2009. 57, 58, 59, 60

[158] Eurosurveilliance. 57, 59

[159] Eurosurveilliance. 57, 59

[160] Reports of the brazilian health department (ministerio da saude). 57, 59

[161] Secretaria de salud, mexico. situation actual de la epidemia, oct 12, 2009. 59

[162] Who, production and availability of pandemic influenza a (h1n1) vaccines. 60

[163] Us food and drugs administration, fda approves vaccines for 2009 h1n1 influenza virus. 60, 61

[164] A. Flahault, E. Vergu, L. Coudeville, and R. Grais. Strategies for containing a global influenza pandemic. *Vaccine*, 24:6751–6755, 2006. 60, 66

[165] N. M. Ferguson, D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437:209, 2005. 60, 66

[166] J.T. Wu, S. Riley, C. Fraser, and G.M. Leung. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med*, 3:e361, 2006. 60, 66

[167] N. Arinaminpathy and A.R. McLean. Antiviral treatment for the control of pandemic influenza: some logistical constraints. *J. R. Soc. Interface*, 5:5945–553, 2008. 60, 66

[168] M.L. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, et al. Mitigation measures for pandemic influenza in italy: An individual based model considering different scenarios. *PLoS ONE*, 3:e1790, 2008. 60, 66

[169] Cdc: Weekly 2009 h1n1 flu media briefing. 60, 61

[170] Us news. 60

[171] Pbs. 60

[172] Lancet. 61

[173] A.C. Singer et al. Meeting report: Risk assessment of tamiflu use under pandemic conditions. *Environ Health Perspect*, 116:1563â1567, 2008. 61

[174] A. Flahault, E. Vergu, and P.-Y. Boelle. Potential for a global dynamic of influenza a(h1n1). *BMC Infect Dis*, 9:129, 2009. 61

[175] Novartis says swine flu shots works with single dose (2009). 61

[176] M.E. Greenberg et al. Response after one dose of a monovalent influenza a (h1n1) 2009 vaccine - preliminary report. *New Engl. J. Med.*, page 10.1056/NEJMoa0907413, 2009. 61

[177] T.W. Clark. Trial of influenza a (h1n1) 2009 monovalent mf59-adjuvanted vaccine - preliminary report. *New Engl. J. Med.*, page 10.1056/NEJMoa0907650, 2009. 61

[178] Cdc advisors make recommendations for use of vaccine against novel h1n1 (2009). 61

[179] Y. Yang et al. The transmissibility and control of pandemic influenza a(h1n1) virus. *Science*, page 10.1126/science.1177373, 2009. 61, 64

[180] G. Chowell et al. Adaptive vaccination strategies to mitigate pandemic influenza: Mexico as a case study. *PLoS Currents Influenza*, page 19:RRN1004., 2009. 61

[181] J. Medlock and A.P. Galvani. Optimizing influenza vaccine distribution. *Science*, 325:1705–1708, 2009. 61

[182] N.E. Basta, M.E. Halloran, L. Matrajt, and I.M. Longini. Estimating influenza vaccine efficacy from challenge and community-based study data. *Am J Epidemiol*, 168:1343â1352, 2008. 61

[183] Ministero della salute. 61

[184] Bundesministerium fÃŒr gesundheit. 61

[185] Ministerio de sanidad y consumo. 61

[186] Ministere de la sante et des sports. 61

[187] Public health agency. 61

[188] J.M. McCaw, J. McVernon, E.S. McBryde, and J.D. Mathews. Influenza: accounting for prior immunity. *Science*, 325:1071, 2009. 66

[189] CDC. *MMWR*, 58:521, 2009. 66

[190] G. Katriel and L. Stone. Pandemic influenza dynamics and the breakdown of herd immunity. *PLoS Currents Influenza*, 2009. 66

[191] M.G. Baker, N. Wilson, Q.S. Huang, S. Paine, L. Lopez, and D. Bandaranayake. Pandemic influenza a(h1n1)v in new zealand: the experience from april to august 2009. *Euro Surveill*, page 14(34):pii=19319, 2009. 66

[192] Uk department of health. swine flu: Uk planning assumptions. issued 3 september, 2009. 67

[193] Reed C, Angulo FJ, Swerdlow DL, Lipsitch M, Meltzer MI, Jernigan D, and et al. Estimates of the prevalence of pandemic (h1n1) 2009, united states, aprilâjuly 2009, 2009. 67

[194] European centre for disease control and prevention, pandemic (h1n1) 2009 daily update (november 23, 2009). 67

[195] Morens DM, Taubenberger, and Fauci AS. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J Infect Dis*, 198:962–970, 2008. 67

[196] Perez-Padilla R, de la Rosa-Zamboni D, Ponce de Leon S, and et al. Pneumonia and respiratory failure from swine-origin influenza a (h1n1) in mexico. *New Engl J Med*, 361:680–689, 2009. 67

[197] CDC. Hospitalized patients with novel influenza a (h1n1) virus infection - california, april-may, 2009. *MMWR*, 58:536–541, 2009. 67

[198] CDC. Intensive-care patients with severe novel influenza a (h1n1) virus infection - michigan, june 2009. *MMWR*, 58:749–752, 2009. 67

[199] J. Rello, A. Rodríguez, P. Ibañez, L. Socias, et al. Intensive care adult patients with severe respiratory failure caused by influenza a (h1n1)v in spain, 2009. 67

[200] CDC. Bacterial coinfections in lung tissue specimens from fatal cases of 2009 pandemic influenza a (h1n1) - united states, may-august 2009. *MMWR*, 58:1–4, 2009. 67

[201] The ANZIC Influenza Investigators. Critical care services and 2009 h1n1 influenza in australia and new zealand. *New Engl J Med*, 361:1925–1934, 2009. 67, 68, 69

[202] World health organization. clinical management of human infection with pandemic (h1n1) 2009: revised guidance, november 2009. 67

[203] W.S. Lim. Pandemic flu: clinical management of patients with an influenza-like illness during an influenza pandemic. *Thorax*, 62:1–46, 2007. 67

[204] Society of critical care medicine, critical care statistics in the united states 2006. 67

[205] The information system of the federal health monitoring. 67

[206] Wunsch H, Angus DC, Harrison DA, and et al. Variation in critical care services across north america and western europe. *Crit Care Med*, 36(10):2787, 2008. 68

[207] Ercole A, Taylor BL, Rhodes A, and Menon DK. Modelling the impact of an influenza a/h1n1 pandemic on critical care demand from early pathogenicity data: the case for sentinel reporting. *Anaesthesia*, 64:937–941, 2009. 68

[208] British thoracic society. 68, 78

[209] National Public Health Service for Wales HPA Northern Ireland Swine influenza investigation teams Health Protection Agency, Health Protection Scotland. Epidemiology of new influenza a (h1n1) virus infection, united kingdom, april - june 2009. *Euro Surveill.*, 14, 2009. 70

[210] Novel influenza A(H1N1) investigation team. Description of the early stage of pandemic (h1n1) 2009 in germany, 27 april-16 june 2009. *Euro Surveill.*, 14, 2009. 70

[211] Influenza a(h1n1) - measures adopted by governments worldwide. 71, 73

[212] B.J. Cowling, L.L.H. Lau, P. Wu, H.W.C. Wong, V. J. Fang, et al. Entry screening to delay local transmission of 2009 pandemic influenza a (h1n1). *BMC Infect. Dis.*, 10:82, 2010. 71

[213] C. Viboud, M.A. Miller, B.T. Grenfell, O.N. Bjornstad, and L. Simonsen. Air travel and the spread of influenza: important caveats. *PLoS Med*, 3:e503, 2006. 71

[214] T.D. Hollingsworth, N.M. Ferguson, and R.M. Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature Med*, 12:497–499, 2006. 71, 76

[215] J.M. Epstein, D.M. Goedecke, F. Yu, R.J. Morris, D.K. Wagener, and G.V. Bobashev. Controlling pandemic flu: The value of international air travel restrictions. *PLoS ONE*, 2:e401, 2007. 71, 76

[216] WHO writing group. Non-pharmaceutical interventions for pandemic influenza, international measures. *Emerg. Infect. Dis.*, 12:81–8, 2006. 71

[217] D.M. Bell. Public health interventions and sars spread, 2003. *Emerg. Infect. Dis.*, 10:1900–1906, 2004. 71

[218] J. Brownstein, C.J. Wolfe, and K.D. Mandl. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the united state. *PLoS Med.*, 3:e401, 2006. 71

[219] Reuters, *argentina confirms first h1n1 flu case, may 7 2009,*. 73

[220] Reuters, *argentina lifting flu-related ban on mexico flights, may 14 2009,*. 73

[221] The san diego union tribune *tijuana-shanghai flights to resume, january 12 2010.* 73

[222] Xinhua news, *china suspends flights from mexico, may 2 2009,*. 73

[223] Ministerio de relaciones exteriores de la república de cuba, *flights between mexico and cuba to be restored from monday, may 29 2009.* 73

[224] Reuters, *peru has its first swine flu case; bans flights, april 30 2009,*. 73

[225] Reuters, *peru lifts ban on mexico flights after flu fears, may 13 2009,*. 73

[226] Reuters, factbox - measures against swine flu in europe, april 30 2009. 73

[227] Reuters, factbox - measures in north, south america against flu, may 4 2009. 73

[228] Reuters, factbox - measures in asia against deadly flu, may 4 2009. 73

[229] Boletín mensual de estadística operacional. secretaría de comunicaciones y transportes. 71

[230] Boletín mensual de estadística operacional de la aviación en méxico. 74

[231] G. Scalia Tomba and J. Wallinga. A simple explanation for the low impact of border control as a countermeasure to the spread of an infectious disease. *Math. Biosci.*, 214:70–72, 2008. 76

[232] A. Gautreau, A. Barrat, and M. Barthélemy. Global disease spread: Statistic and estimation on arrival times. *J. Theo. Bio.*, 251:509–522, 2008. 76

[233] V. Colizza and A. Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251:450–467, 2008. 76, 79, 81

[234] year=2000 C. L. Barrett *et al.*, journal=Technical Report LA-UR-00-1725, Los Alamos National Laboratory. 76, 78

[235] S. P. Gorman P. Nijkamp F.-J. Bade. R. Patuelli, A. Reggiani. 76, 78

[236] R. Guimerá, S. Mossa, A. Turtschi, and L.A.N. Amaral. 76

[237] Italian national institute for statistics (istat). 78

[238] Harris, T.E. *The theory of branching processes.* Dover Publications, 1989. 79

[239] A. Vazquez. Polynomial growth in branching processes with diverging reproductive number. *Phys. Rev. Lett.*, 96:038702, 2006. 79

[240] V. Colizza and A. Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.*, 99:148701, 2007. 79, 81

[241] Murray, J.D. *Mathematical Biology.* 3rd edition Berlin: Springer Verla, 2005. 80, 82

[242] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of Networks. *Advances in Physics*, **51**:1079–1187, 2002. 81

[243] E. Brooks-Pollock, W.J. Edmunds, and K.T.D. Eames. Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 h1n1v influenza epidemic in england. *BMC Inf Dis*, 11:68, 2011. 85, 86

[244] World Health Organization. Seroepidemiological studies of pandemic influenza a(h1n1) 2009 virus. *Weekly Epidemiological Record*, 85:24, 2010. 86

[245] F. Carrat, C. Pelat, D. Levy-Bruhl, I. Bonmarin, and N. Lapidus. Planning for the next influenza h1n1 season: a modelling study. *BMC Inf Dis*, 10:301, 2010. 86

[246] D.M. Fleming. Influenza surveillance, the swine-flu pandemic, and the importance of virology. *Clinical Evidence*, 2009. 86

[247] A.J. van Hoek and E. Miller. Response to guest editorial "influenza surveillance, the swine-flu pandemic, and the importance of virology". *Clinical Evidence*, 2010. 86

[248] Food UK Department for Environment and Rural Affairs (DEFRA). Origin of the uk foot and mouth disease epidemic 2001. 2002. 89

[249] L.H. Taylor, S.M. Latham, and M.E. Woolhouse. Risk factors for human disease emergence. *Phil Trans R Soc B*, 356:983–989, 2001. 89

[250] I. Anderson. Foot mouth disease 2001: Lessons to be learned inquiry report. *London, UK: The Stationary Office*, 2002. 89

[251] European Parliament and European Council. Regulation (ec) no. 1760/2000 of 17 july 2000 establishing a system for the identification and registration of bovine animals and regarding labeling of beef and beef products and repealing council regulation (ec) no. 820/97 european council. off. j. eur. communities l 204, 1-10, 2000. 89

[252] MAF Biosecurity New Zealand. Review of selected cattle identification and tracing systems worldwide, 2009. 89

[253] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45, 2003. 89, 95

[254] C. Dube, C. Ribble, D. Kelton, and B. McNab. A review of networks analysis terminology and its application to foot-and-mouth disease modeling and policy development. *Transboundary and Emerging Diseases*, 56:73–85, 2009. 89

[255] B. Martinez-Lopez, A.M. Perez, and J.M. Sanchez-Vizcaino. Social network analysis. review of general concepts and use in preventive veterinary medicine. *Transboundary and Emerging Diseases*, 56:109–120, 2009. 89

[256] F. Natale, A. Giovannini, L. Savini, D. Palma, L. Possenti, et al. Network analysis of italian cattle trade patterns and evaluation of risks for potential disease spread. *Prev Vet Med*, 92:341–350, 2009. 90, 91, 95, 118, 119

[257] M. Bigras-Poulin, R.A. Thompson, M. Chriel, S. Mortensen, and M. Greiner. Network analysis of danish cattle industry trade patterns as an evaluation of risk potential for disease spread. *Prev Vet Med*, 76:11–39, 2006. 90, 92, 95, 104

[258] S. Rautureau, B. Dufor, and B. Durand. Vulnerability of animal trade networks to the spread of infectious diseases: a methodological approach applied to evaluation and emergency control strategies in cattle, france, 2005. *Transboundary and Emerging Diseases*, 2010. 90, 95, 97, 101, 118, 119

[259] D.U. Ortiz-Pelaez, Pfeiffer, R.J. Soares-Magalhaes, and F.J. Guitian. Use of social network analysis to charaterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (fmd) epidemic in the uk. *Prev Vet Med*, 76:40–55, 2006. 90, 95, 118, 119

[260] V.V. Volkova, R. Howey, N.J. Savill, and M.E.J. Woolhouse. Potential for transmission of infections in networks of cattle farms. *Epidemics 2*, pages 116–122, 2010. 90, 91, 95, 97, 102, 118, 119

[261] R.M. Christley, S.E. Robinson, R. Lysons, and N.P. French. Network analysis of cattle movement in great britain. *Proc. Soc. Vet. Epidemiol. Prev. Med*, pages 234–243, 2005. 90, 91, 92, 95, 102, 118, 119

[262] R.R. Kao, L. Danon, D.M. Green, and I.Z. Kiss. Demographic structure and pathogen dynamics on the network of livestock movements in great britain. *Proc R Soc Lond B Biol Sci*, 273:1999–2007, 2006. 90, 95

[263] M.E. Woolhouse, D.J. Shaw, L. Matthews, W.C. Liu, D.J. Mellor, et al. Epidemiological implications of the contact network structure for cattle farms and the 20-80 rule. *Biol Lett*, 1:350–352, 2005. 90, 95

[264] M.L. Brennan, R. Kemp, and R.M. Christley. Direct and indirect contacts between cattle farms in north-west england. *Prev Vet Med*, 84:242–260, 2008. 90, 95

[265] S.E. Robinson, M.G. Everett, and R.M. Christley. Recent network evolution increases the potential for large epidemics in the british cattle population. *J. R. Soc. Interface*, 4:669–674, 2007. 90, 95, 118, 119

[266] M.C. Vernon and M.J. Keeling. Representing the uk's cattle herd as static and dynamic networks. *Proc R Soc B*, 276:469–476, 2009. 90, 97, 117, 120

[267] S.E. Robinson and R.M. Christley. Exploring the role of auction markets in cattle movements within great britain. *Prev Vet Med*, 14:21–37, 2007. 90, 118, 119

[268] Lloyd A.L. and R.M. May. How viruses spread among computers and people. *Science*, 292:1316–1317, 2001. 90

[269] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *J Theor Biol*, 235:275–288, 2005. 90

[270] Vázquez A., B. Rácz, A. Lukács, and A.L. Barabási. Impact of non-poissonian activity patterns on spreading processes. *Phys Rev Lett*, 98:158702, 2007. 90

[271] J. Gómez-Gardenes, M. Campillo, L.M. Floría, and Y. Moreno. Dynamical organization of cooperation in complex topologies. *Phys Rev Lett*, 98:108103, 2007. 90

[272] M. Perc. Evolution of cooperation on scale-free networks subject to error and attack. *New J. Phys.*, 11:033027, 2007. 90

[273] A. Szolnoki, M. Perc, and Z. Danku. Making new connections towards cooperation in the prisoner's dilemma game. *EPL*, 84:50007, 2008. 90

[274] B. Guerra, J. Poncela, J. Gómez-Gardenes, V. Latora, and Y. Moreno. Dynamical organization towards consensus in the axelrod model on complex networks. *Phys. Rev. E*, 81:056105, 2010. 90

[275] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591, 2009. 90

[276] C. Castellano, D. Vilone, and A. Vespignani. Incomplete ordering of the voter model on small-world networks. *Europhys. Lett.*, 63:153–158, 2003. 90

[277] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: structure and dynamics. *Physics Reports*, 424, 2006. 90

[278] K. Klemm, V.M. Eguíluz, R. Toral, and M. San Miguel. Nonequilibrium transitions in complex networks: a model of social interaction. *Phys. Rev. E*, 67:026120, 2003. 90

[279] Italian national animal identification and registration database instituted by ministry of health at csn of istituto g.caporale of teramo. 90

[280] M.J. Keeling, L. Danon, M.C. Vernon, and T.A. House. Individual identity and movement networks for disease metapopulations. *Proc Natl Acad Sci USA*, 107:8866–8870, 2010. 92, 97

[281] V. Latora and M. Marchiori. Is the boston subway a small-world network? *Physica A*, 314:109–113, 2001. 93

[282] C. Roth, S.M. Kang, M. Batty, and M. Barthélemy. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE*, page 6(1):e15923, 2011. 93

[283] M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, et al. Scaling behavior in the growth of companies. *Nature*, 379:804–806, 1996. 93, 109

[284] D. Braha and Y. Bar-Yam. From centrality to temporary fame: Dynamic centrality in complex networks. *Complexity*, 12:59–63, 2006. 109

[285] M.A. Serrano, M. Boguna, and A. Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci USA*, 106:6483–6488, 2009. 111

[286] J. Stehlé, A. Barrat, and G. Bianconi. Dynamical and bursty interactions in social networks. *Phys. Rev. E*, 81:035101, 2010. 111

[287] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, et al. Network motifs: simple building blocks of complex networks. *Science*, 298:824–827, 2002. 113

[288] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, et al. What's in a crowd? analysis of face-to-face behavioral networks. *J Theor Biol*, 271:166–180, 2011. 113, 117

[289] V. Kostakos. Temporal graphs. *Physica A*, 388:1007–1023, 2009. 113

[290] K. Lerman, R. Ghosh, and J.H. Kang. Centrality metrics for dynamic networks. *Proc. of the 8th workshop on Mining and Learning with graphs (ACM, New York)*, 2010. 113

[291] R.V. Pan and J. Saramaki. Path lengths, correlations, and centrality in temporal networks. *arXiv:1101.5913*, 2011. 113

[292] L. Kovanen, M. Karsai, K. Kaski, J. Kertész, and J. Saramäki. Temporal motifs in time-dependent networks. *arXiv:1107.5646*, 2011. 113

[293] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A*, 333:529–540, 2004. 114

[294] S.S. Morse. Factors in the emergence of infectious diseases. *Emerg Infect Dis*, 1:7–15, 1995. 116

[295] L.E.C. Rocha, F. Liljeros, and P. Holme. Information dynamics shape the sexual networks of internet-mediated prostitution. *Proc. Natl. Acad. Sci USA*, 107:5706–5711, 2009. 117

[296] M. Karsai, M. Kivelä, R.K. Pan, K. Kaski, J. Kertész, et al. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E*, 83:025102, 2010. 117

[297] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza. Dynamical patterns of cattle trade movements. *PLoS ONE*, 6(5):e19869, 2011. 117, 127, 133

[298] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378, 2000. 117

[299] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdown. *Phys. Rev. Lett.*, 85:4646, 2000. 117

[300] P. Holme, J. Kim, C.N. Yoon, and S.K. Han. Attack vulnerability of complex networks. *Phys Rev E*, 65:056109, 2002. 117

[301] L. Dall'Asta, A. Barrat, M. Barthélemy, and A. Vespignani. Vulnerability of weighted networks. *J Stat Mech*, page P04006, 2006. 117

[302] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proc Natl Acad Sci USA*, 101:2658–2663, 2004. 117

[303] F. Natale, L. Savini, A. Giovannini, P. Calistri, L. Candeloro, et al. Evaluation of risk and vulnerability using a disease flow centrality measure in dynamic cattle trade networks. *Prev Vet Med*, 98:111–118, 2011. 118

[304] R. Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.*, 15:237–240, 1969. 119

[305] M.J. Keeling. Models of foot-and-mouth disease. *Proc. R Soc. B*, 272:1195–1202, 2005. 119

[306] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Brockmann. The structure of borders in a small world. *PLoS ONE*, 5(11): e15422, 2010. 122